THE ANNALS *of* MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. C. CARVER, 1930-1938
EDITED BY S. S. WILKS, 1938-1949
THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

PAGE

Saddlepoint Approximations in Statistics. H. E. DANIELS 631
A General Theory of Discrimination When the Information About Alternative Population Distributions is Based on Samples. C. RADHAKRISHNA RAO 651
A Confidence Interval for Variance Components. J. R. GREEN 671
A Simple Sequential Procedure for Testing Statistical Hypotheses. CHIA KUEI TSAO 687
On a Contagious Distribution. R. S. G. RUTHERFORD 703
Identification and Estimation of Linear Manifolds in n-Dimensions. T. A. JEEVES 714
Asymptotic Behavior of Some Rank Tests for Analysis of Variance. FRED C. ANDREWS 724
Multidimensional Stochastic Approximation Methods. JULIUS R. BLUM 737
Certain Inequalities in Information Theory and the Cramer-Rao Inequality. S. KULLBACK 745
Some Further Results in Simultaneous Confidence Interval Estimation. S. N. ROY 752
Tables for the Distribution of the Number of Exceedances. BENJAMIN EPSTEIN 762
On the Factorisation of Distributions. HENRY TEICHER 769
On the Convolution of Distributions. HENRY TEICHER 775
Notes:
Sequential Procedures that Control the Individual Probabilities of Coming to the Various Decisions. LIONAL WEISS 779
On a Characterisation of the Gamma Distribution. R. G. LAHA 784
Matching in Paired Comparisons. J. L. HODGES, JR. and E. L. LEHMANN. 787
Note on a Theorem of Lionel Weiss. LUCIEN LECAM 791
The Distribution of Distance in a Hypersphere. R. D. LORD 794
Extreme Values in Samples from m-Dependent Stationary Stochastic Processes. G. S. WATSON 798
Expression of the k-Statistics, k_9 and k_{10} in Terms of Power Sums and Sample Moments. M. ZIA UD-DIN 800
The Probability Integral of Range for Samples from a Symmetrical Unimodal Population. J. H. CADWELL 803
Abstracts of Papers 806
News and Notices 814
Report of the Montreal Meeting 820
Publications Received 826
Institutional Members 828

Vol. 25, No. 4 — December, 1954

THE [illegible]
OF MATHEM[illegible]ATISTIC[illegible]

[illegible]

DAVID BLACKWELL [illegible] WA[illegible] [illegible]ANDER M. [illegible]
[illegible], JR. WILL[illegible] [illegible]

WITH [illegible]

T. E[illegible] [illegible] Pa[illegible]
[illegible]
J. KIEFER [illegible]
WILLIAM [illegible]
[illegible]
[illegible] H[illegible]
G. E. NOE[illegible] [illegible]M. S[illegible]
[illegible] WALD [illegible]

[illegible] quarterly by the Inst[illegible] [illegible]atistics in Ma[illegible]
June, September and D[illegible]

INSTITUTE OF MATHEMATICAL STATISTICS

[illegible] Department of Mathematics, [illegible], East Lansing, Michigan
H. J. Arnold, Secretary-Treasurer
This address should [illegible] communications concerning membership, [illegible] of address, back numbers, etc., but [illegible]. Changes in [illegible] [illegible] a given issue [illegible] on or before the 15th of the month preceding [illegible]

[illegible] Statistical Laboratory, [illegible], Berkeley 4, California
E. L. Lehmann, [illegible]
Manuscripts should be submitted to this address; each manuscript should be typewritten, double-spaced with wide [illegible] and the original copy [illegible] submitted (preferably with one additional copy). [illegible] acceptable if they [illegible] printer [illegible] will take [illegible] minimum and [illegible] at the end of the pa[illegible] be avoided. Figures, charts, [illegible] drawn on plain white pa[illegible] ink [illegible] the size they [illegible] to [illegible] in m[illegible] [illegible] reprin[illegible]

[illegible] $12 per year [illegible] issues $3.50. [illegible] $3.50 per issue. [illegible]

[illegible]
WAVERLY [illegible] Baltimore, [illegible]

[illegible]

SADDLEPOINT APPROXIMATIONS IN STATISTICS[1]

By H. E. Daniels

University of Cambridge and University of Chicago

1. Introduction and summary. It is often required to approximate to the distribution of some statistic whose exact distribution cannot be conveniently obtained. When the first few moments are known, a common procedure is to fit a law of the Pearson or Edgeworth type having the same moments as far as they are given. Both these methods are often satisfactory in practice, but have the drawback that errors in the "tail" regions of the distribution are sometimes comparable with the frequencies themselves. The Edgeworth approximation in particular notoriously can assume negative values in such regions.

The characteristic function of the statistic may be known, and the difficulty is then the analytical one of inverting a Fourier transform explicitly. In this paper we show that for a statistic such as the mean of a sample of size n, or the ratio of two such means, a satisfactory approximation to its probability density, when it exists, can be obtained nearly always by the method of steepest descents. This gives an asymptotic expansion in powers of n^{-1} whose dominant term, called the saddlepoint approximation, has a number of desirable features. The error incurred by its use is $O(n^{-1})$ as against the more usual $O(n^{-1/2})$ associated with the normal approximation. Moreover it is shown that in an important class of cases the *relative* error of the approximation is uniformly $O(n^{-1})$ over the whole admissible range of the variable.

The method of steepest descents was first used systematically by Debye for Bessel functions of large order (Watson [17]) and was introduced by Darwin and Fowler (Fowler [9]) into statistical mechanics, where it has remained an indispensable tool. Apart from the work of Jeffreys [12] and occasional isolated applications by other writers (e.g. Cox [2]), the technique has been largely ignored by writers on statistical theory.

In the present paper, distributions having probability densities are discussed first, the saddlepoint approximation and its associated asymptotic expansion being obtained for the probability density of the mean $\bar{x}$ of a sample of n. It is shown how the steepest descents technique is related to an alternative method used by Khinchin [14] and, in a slightly different context, by Cramér [5]. General conditions are established under which the relative error of the saddlepoint approximation is $O(n^{-1})$ uniformly for all admissible $\bar{x}$, with a corresponding result for the asymptotic expansion. The case of discrete variables is briefly discussed, and finally the method is used for approximating to the distribution of ratios.

2. Mean of n independent identically distributed random variables. Let x be a continuously distributed random variable with distribution function $F(x)$.

Received 1/16/53, revised 3/31/54.

[1] Research carried out partly under sponsorship of the Office of Naval Research.

Assume that a density function $f(x) = F'(x)$ exists and suppose the moment-generating function

$$M(T) = e^{K(T)} = \int_{-\infty}^{\infty} e^{Tx} f(x)\, dx$$

converges for real T in some nonvanishing interval containing the origin. Let $-c_1 < T < c_2$ be the largest such interval, where $0 \leqq c_1 \leqq \infty$ and $0 \leqq c_2 \leqq \infty$ but $c_1 + c_2 > 0$. Thus either c_1 or c_2 may be zero, though not both, and the moments need not all exist.

Consider the mean $\bar{x}$ of n independent x's. Its density function $f_n(\bar{x}) = F_n'(\bar{x})$ is given by the usual Fourier inversion formula

$$f_n(\bar{x}) = \frac{n}{2\pi} \int_{-\infty}^{\infty} M^n(it) e^{-nit\bar{x}}\, dt \tag{2.1}$$

(More generally $\int_{-\infty}^{\infty}$ may be replaced by $\lim_{t\to\infty} \int_{-t}^{t}$, but the argument is unaffected.) It is convenient here to employ the equivalent inversion formula

$$f_n(\bar{x}) = \frac{n}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} e^{n[K(T)-T\bar{x}]}\, dT \tag{2.2}$$

where $-c_1 < \Re(T) < c_2$ on the path of integration, and $K(T)$ is the cumulant-generating function.

When n is large, an approximation to $f_n(\bar{x})$ is found by choosing the path of integration to pass through a saddlepoint of the integrand in such a way that the integrand is negligible outside its immediate neighbourhood. The saddlepoints are situated where the exponent has zero derivative, that is where

$$K'(T) = \bar{x}. \tag{2.3}$$

We shall prove in Section 5 that under general conditions (2.3) has a single real root T_0 in $(-c_1, c_2)$ for every value of $\bar{x}$ such that $0 < F_n(\bar{x}) < 1$, and that $K''(T_0) > 0$. Let us choose the path of integration to be a straight line through T_0 parallel to the imaginary axis. Since $K(T) - T\bar{x}$ has a minimum at T_0 for real T, the modulus of the integrand must have a maximum at T_0 on the chosen path. Now we can show by a familiar argument (cf. Wintner [18], p. 14) that on *any* admissible straight line parallel to the imaginary axis the integrand attains its maximum modulus only where the line crosses the real axis. For on the line $T = \tau + iy$,

$$\begin{aligned} |M(T)e^{-T\bar{x}}| &= e^{-\tau\bar{x}} \left| \int_{-\infty}^{\infty} e^{(\tau+iy)x}\, dF(x) \right| \\ &\leqq e^{-\tau\bar{x}} M(\tau). \end{aligned}$$

Equality cannot hold for some $y \neq 0$, otherwise $\int_{-\infty}^{\infty} e^{(\tau+iy)x}\, dF(x) = M(\tau)e^{i\alpha}$

so that $\int_{-\infty}^{\infty} e^{\tau x}[1 - \cos(yx - \alpha)]\, dF(x) = 0$, which contradicts the existence of a density function. Moreover, since $M(\tau + iy) = 0(|y|^{-1})$ for large $|y|$ by the Riemann Lebesgue lemma, the integrand cannot approach arbitrarily near its maximum modulus as $|y|$ becomes large. Consequently, for the particular path chosen, only the nieghbourhood of T_0 need be considered when n is large.

The argument then proceeds formally as follows. On the contour near T_0,

$$K(T) - T\bar{x} = K(T_0) - T_0\bar{x} - \tfrac{1}{2}K''(T_0)y^2 - \tfrac{1}{6}K'''(T_0)iy^3 + \tfrac{1}{24}\cdot K^{iv}(T_0)y^4 + \cdots. \tag{2.4}$$

Setting $y = v/[nK''(T_0)]^{1/2}$ and expanding the integrand we get

$$f_n(\bar{x}) \sim \frac{1}{2\pi}\left[\frac{n}{K''(T_0)}\right]^{1/2} e^{n[K(T_0)-T_0\bar{x}]} \cdot \int_{-\infty}^{\infty} e^{-v^2/2}\left\{1 - \tfrac{1}{6}\lambda_3(T_0)\frac{iv^3}{n^{1/2}} + \frac{1}{n}[\tfrac{1}{24}\lambda_4(T_0)v^4 - \tfrac{1}{72}\lambda_3^2(T_0)v^6] + \cdots\right\} dv \tag{2.5}$$

where $\lambda_j(T) = K^{(j)}(T)/[K''(T)]^{j/2}$ for $j \geqq 3$. The odd powers of v vanish on integration and we obtain an expansion in powers of n^{-1},

$$f_n(\bar{x}) \sim g_n(\bar{x})\left\{1 + \frac{1}{n}[\tfrac{1}{8}\lambda_4(T_0) - \tfrac{5}{24}\lambda_3^2(T_0)] + \cdots\right\} \tag{2.6}$$

where $g_n(\bar{x}) = [n/2\pi K''(T_0)]^{1/2} e^{n[K(T_0)-T_0\bar{x}]}$. We call $g_n(\bar{x})$ the *saddlepoint approximation* to $f_n(\bar{x})$.

3. The method of steepest descents. It is not apparent from the above formal development that (2.6) is a proper asymptotic expansion in which the remainder is of the same order as the last term neglected. The asymptotic nature of an expansion of this type is usually established by the method of steepest descents with the aid of a lemma due to Watson [17], the path of integration being the curve of steepest descent through T_0, upon which the modulus of the integrand decreases most rapidly. An account of the method is given by Jeffreys and Jeffreys [13]. The analysis is simplified by using a "truncated" version of Watson's lemma introduced by Jeffreys and Jeffreys for this purpose.[2] The special form appropriate to the present discussion is as follows.

LEMMA. *If $\psi(z)$ is analytic in a neighbourhood of $z = 0$ and bounded for real $z = w$ in an interval $-A \leqq w \leqq B$ with $A > 0$ and $B > 0$, then*

$$\left(\frac{n}{2\pi}\right)^{1/2}\int_{-A}^{B} e^{-nw^2/2}\psi(w)\, dw \sim \psi(0) + \frac{1}{2n}\psi''(0) + \cdots + \frac{1}{(2n)^r}\frac{\psi^{(2r)}(0)}{r!} + \cdots \tag{3.1}$$

is an asymptotic expansion in powers of n^{-1}.

[2] The proof given in [13] contains an error which will be corrected in the forthcoming new edition.

To apply the lemma, deform the contour so that for $|T - T_0| \leqq \delta$ the line $T = T_0 + iy$ is replaced by the curve of steepest descent which is that branch of $\mathcal{I}\{K(T) - T\bar{x}\} = 0$ touching $T = T_0 + iy$ at T_0, when δ is chosen small enough to exclude possible saddlepoints other than T_0. The contour is thereafter continued along the orthogonal curves of constant $\mathcal{R}\{K(T) - T\bar{x}\}$. These can easily be shown to meet the original path in points $T_0 - i\alpha$ and $T_0 + i\beta$ where $\alpha > 0$ and $\beta > 0$, if δ is small enough, since T_0 is a simple root of (2.3). The rest of the contour remains as before.

On the steepest descent curve, $K(T) - T\bar{x}$ is real and decreases steadily on each side of T_0. Make the substitution

$$\begin{aligned} -\tfrac{1}{2}w^2 &= K(T) - T\bar{x} - K(T_0) + T_0\bar{x} \\ &= \tfrac{1}{2}K''(T_0)(T - T_0)^2 + \tfrac{1}{6}K'''(T_0)(T - T_0)^3 + \cdots \qquad (3.2) \\ &= \tfrac{1}{2}z^2 + \tfrac{1}{6}\lambda_3(T_0)z^3 + \tfrac{1}{24}\cdot\lambda_4(T_0)z^4 + \cdots, \end{aligned}$$

where $z = (T - T_0)[K''(T_0)]^{1/2}$, and w is chosen to have the same sign as $\mathcal{I}(z)$ on the contour. Inversion of the series yields an expansion

$$z = iw + \tfrac{1}{6}\lambda_3(T_0)w^2 + \{\tfrac{1}{24}\cdot\lambda_4(T_0) - \tfrac{5}{72}\lambda_3^2(T_0)\}iw^3 + \cdots$$

convergent in some neighbourhood of $w = 0$. The contribution to (2.2) from this part of the contour is then

$$\frac{n}{2\pi i}\frac{e^{n[K(T_0)-T_0\bar{x}]}}{[K''(T_0)]^{1/2}}\int_{-A}^{B} e^{-nw^2/2}\frac{dz}{dw}\,dw,$$

to which Watson's lemma can be applied. Contributions to the integral from the rest of the contour are negligible since for $T = T_0 + iy$ with y outside $(-\alpha_1, \alpha_2)$ we have

$$|M(T)e^{-T\bar{x}}| \leqq \rho\,|M(T_0)e^{-T_0\bar{x}}|$$

for some $\rho < 1$, so that the extra terms contain the factor ρ^n and may be neglected. We thus obtain the asymptotic expansion

$$f_n(\bar{x}) \sim \left[\frac{n}{2\pi K''(T_0)}\right]^{1/2} e^{n[K(T_0)-T_0\bar{x}]}\left\{a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots\right\}. \qquad (3.3)$$

From the Lagrange expansion of dz/dw we find

$$a_r = \frac{1}{2^r r!}\frac{d^{2r}}{dz^{2r}}\left\{\frac{z}{iw(z)}\right\}^{2r+1}\Bigg|_{z=0}. \qquad (3.4)$$

The coefficients of this series can be shown to be identical with those obtained by the method of Section 2 (see Appendix).

4. A generalisation of the Edgeworth expansion. We now show how the work of Cramér [3], [4] on the Edgeworth series can also be employed to establish the

asymptotic nature of (2.6), using a technique similar to that adopted by Cramér [5] and Khinchin [14].

It has been proved that on any admissible path of the form $T = \tau + iy$ the integrand attains its maximum modulus only at $T = \tau$. Consequently (2.6) is only one of a family of series for $f_n(\bar{x})$ which can be derived in a similar way by integrating along $T = \tau + iy$, τ taking any value in $(-c_1, c_2)$. In particular, $\tau = 0$ gives the Edgeworth series, whose asymptotic character was demonstrated by Cramér (3).

We have

$$e^{K(T)-T\bar{x}} = \int_{-\infty}^{\infty} e^{T(x-\bar{x})} f(x)\,dx = \int_{-\infty}^{\infty} e^{Tu} f(u+\bar{x})\,du. \tag{4.1}$$

On the path $T = \tau + iy$ we can put

$$e^{K(T)-T\bar{x}} = e^{K(t)-\tau\bar{x}}\phi(y),$$

where

$$\phi(y) = \frac{\int_{-\infty}^{\infty} e^{iyu}\cdot e^{\tau u} f(u+\bar{x})\,du}{\int_{-\infty}^{\infty} e^{\tau u} f(u+\bar{x})\,du} \tag{4.2}$$

is the characteristic function for a random variable u having the density function $h(u) \propto e^{\tau u} f(u + \bar{x})$. The inversion formula (2.2) then becomes

$$\begin{aligned} f_n(\bar{x}) &= e^{n[K(\tau)-t\bar{x}]}\cdot(n/2\pi)\int_{-\infty}^{\infty} \phi^n(y)\,dy \\ &= e^{n[K(\tau)-\tau\bar{x}]} h_n(0) \end{aligned}$$

where $h_n(\bar{u})$ is the density function for the mean $\bar{u}$ of n independent u's. Using the fact that

$$\log \phi = [K'(\tau) - \bar{x}]iy + \sum_{j\geq 2} K^{(j)}(\tau)\frac{(iy)^j}{j!}$$

we may replace $h_n(0)$ by its Edgeworth series and obtain the family of asymptotic expansions

$$\begin{aligned} f_n(\bar{x}) \sim \exp n\{K(\tau) - \tau\bar{x} - [K'(\tau) - \bar{x}]^2/2K''(\tau)\} \\ \cdot[n/2\pi K''(\tau)]^{1/2}\{1 + A_1/n^{1/2} + A_2/n + \cdots\} \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} A_1 &= (1/3!)\lambda_3(\tau)\, H_3([K'(\tau) - \bar{x}][n/K''(\tau)]^{1/2}), \\ A_2 &= (1/4!)\lambda_4(\tau)\, H_4([K'(\tau) - \bar{x}][n/K''(\tau)]^{1/2}) \\ &\quad + (10/6!)\lambda_3^2(\tau)\, H_6([K'(\tau) - \bar{x}][n/K''(\tau)]^{1/2}), \end{aligned}$$

etc., the H's being Hermite polynomials.

When $\tau = 0$ this reduces to the Edgeworth series for $f_n(\bar{x})$. (Since c_1 or c_2 can be zero it may not be possible to take the expansion beyond a certain number of terms in this case). On the other hand when $\tau = T_0$, so that $K'(T_0) = \bar{x}$, all the odd powers of $n^{-1/2}$ vanish and we get (2.6), which is an expansion in powers of n^{-1}. In particular the dominant term $g_n(\bar{x})$ has the same accuracy as the first two terms of the Edgeworth series. Unlike the latter, however, $g_n(\bar{x})$ can never be negative, and is shown in Section 7 to have a further important advantage over the other approximations.

5. Examples. The method is applied to three examples.

EXAMPLE 5.1.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}, \qquad -\infty \leqq x \leqq \infty.$$

$$K(T) = mT + \tfrac{1}{2}\sigma^2 T^2, \qquad K'(T) = m + \sigma^2 T_0 = \bar{x},$$

$$T_0 = (\bar{x} - m)/\sigma^2, \qquad K''(T_0) = \sigma^2,$$

$$g_n(\bar{x}) = \frac{1}{\sigma}\left(\frac{n}{2\pi}\right)^{1/2} e^{-n(\bar{x}-m)^2/2\sigma^2}.$$

In this case $g_n(\bar{x}) = f_n(\bar{x})$ for every value of n.

EXAMPLE 5.2. $\qquad f(x) = (c^\alpha/\Gamma(\alpha))x^{\alpha-1}e^{-cx}, \qquad 0 \leq x \leq \infty.$

$$K(T) = -\alpha \log(1 - T/c), \qquad K'(T_0) = \alpha/(c - T_0) = \bar{x},$$

$$K''(T_0) = \alpha/(c - T_0)^2 = \bar{x}^2/\alpha,$$

$$g_n(\bar{x}) = (n\alpha/2\pi)^{1/2} e^{n\alpha}(c/\alpha)^{n\alpha}\, \bar{x}^{n\alpha-1}\, e^{-nc\bar{x}}.$$

The exact result is

$$f_n(\bar{x}) = [(nc)^{n\alpha}/\Gamma(n\alpha)]\bar{x}^{n\alpha-1}e^{-nc\bar{x}}$$

which differs from $g_n(\bar{x})$ only in that $\Gamma(n\alpha)$ is replaced by Stirling's approximation in the normalising factor. As this can always be readjusted ultimately to make the total probability unity, we can regard $g_n(\bar{x})$ as being in this sense "exact" for all n.

EXAMPLE 5.3. $\qquad f(x) = \tfrac{1}{2}, \qquad -1 \leqq x \leqq 1.$

The density function for the mean of n independent rectangular variables in $(-1, 1)$ is known to be

$$f_n(\bar{x}) = \frac{n^n}{2^n(n-1)!}\sum_{s=0}^{n}(-1)^s\binom{n}{s}\left\langle 1 - \bar{x} - \frac{2s}{n}\right\rangle^{n-1}, \qquad |\bar{x}| \leqq 1$$

where $\langle z\rangle = z$ for $z \geqq 0$ and $=0$ for $z < 0$. (Seal [16] gives a historical note on this result.) We have

$$K(T) = \log\left(\frac{\sinh T}{T}\right), \qquad K'(T_0) = \coth T_0 - \frac{1}{T_0} = \bar{x},$$

$$K''(T_0) = \frac{1}{T_0^2} - \operatorname{cosech}^2 T_0,$$

$$g_n(\bar{x}) = \left(\frac{n}{2\pi}\right)^{1/2} \left\{\frac{1}{T_0} - \operatorname{cosech}^2 T_0\right\}^{-1/2} \left(\frac{\sinh T_0}{T_0}\right)^n e^{-T_0 \bar{x}}$$

When T_0 is large and positive, $\bar{x} \sim 1 - 1/T_0$ and

$$K(T_0) \sim \log (e^{T_0}/2T_0), \qquad K''(T_0) \sim 1/T_0^2.$$

So for small $1 - \bar{x}$,

$$g_n(\bar{x}) \sim (n/2\pi)^{1/2}(\tfrac{1}{2}e)^n(1 - \bar{x})^{n-1}$$

which agrees with $f_n(\bar{x}) = [n^n/2^n(n-1)!](1-\bar{x})^{n-1}$ when $\bar{x} > 1 - 2/n$ except for the normalising constant, and there is similar agreement for $\bar{x}$ near -1. Actually $\log_e g_n(\bar{x})$ is remarkably close to $\log_e f_n(\bar{x})$ for quite moderate values of n over the whole range of x. Table 1 shows the agreement for $n = 6$, which could be improved by adjusting the normalising constant. With n as low as 6, $g_n(\bar{x})$ never differs from $f_n(\bar{x})$ by as much as 4 per cent. This example leads one to

TABLE 1

$\bar{x}$	.1	.2	.3	.4	.5	.6	.7	.8	.9
$\log_e f_6(\bar{x})$	0.419	0.172	−0.249	−0.860	−1.687	−2.778	−4.216	−6.243	−9.709
$\log_e g_6(\bar{x})$	0.445	0.199	−0.221	−0.829	−1.653	−2.742	−4.188	−6.228	−9.695
Difference	0.026	0.027	0.028	0.031	0.034	0.036	0.028	0.015	0.014

enquire under what conditions $f_n(\bar{x})/g_n(\bar{x}) \to 1$ *uniformly for all* $\bar{x}$ as $n \to \infty$, so that the relative accuracy of the approximation is maintained up to the ends of the range of $\bar{x}$. In Section 7 we show that the result is true for a wide class of density functions.

6. The real roots of $K'(T) = \xi$. In this section we discuss the existence and properties of the real roots of the equation $K'(T) = \xi$, upon which the approximation $g_n(\bar{x})$ depends. The conditions are here relaxed so that the distribution need not have a density function. The moment generating function is still assumed to satisfy the conditions of Section 2, namely that

$$M(T) = e^{K(T)} = \int_{-\infty}^{\infty} e^{Tx}\, dF(x)$$

converges for real T in $-c_1 < T < c_2$ where $0 \leqq c_1 \leqq \infty$ and $0 \leqq c_2 \leqq \infty$ but $c_1 + c_2 > 0$. Throughout this section T is supposed to take real values only.

The distribution may extend from $-\infty$ to ∞, or it may be limited at either or both ends. We shall write

$$
\begin{aligned}
F(x) &= 0, && x < a,\\
0 < F(x) &< 1, && a < x < b,\\
F(x) &= 1, && b < x,
\end{aligned}
$$

where if desired $a = -\infty$ or $b = \infty$, or both. Note that $b < \infty$ implies $c_2 = \infty$ so that $c_2 < \infty$ implies $b = \infty$, and similarly for a and c_1. The converse is not true since b and c_2 (or a and c_1) can both be infinite.

We now establish the conditions under which $K'(T) = \xi$ has no real root when ξ lies outside the interval (a, b), and has a unique simple root T_0 for every ξ in (a, b). It is convenient to consider first the case where both a and b are finite.

THEOREM 6.1. *$F(x) = 0$ for $x < a$, and $F(x) = 1$ for $x > b$ if and only if $K(T)$ exists for all real T and $K'(T) = \xi$ has no real root whenever $\xi < a$ or $\xi > b$.*

PROOF. Write

$$M(T, \xi) = e^{K(T)-T\xi} = \int_{-\infty}^{\infty} e^{T(x-\xi)}\, dF(x).$$

If $dF(x) = 0$ outside (a, b) then $M(T, \xi)$ exists for all real T and

$$M'(T,\xi) = \int_a^b (x - \xi)e^{T(x-\xi)}\, dF(x)$$

exists and has constant sign for all T when $\xi < a$ or $\xi > b$, and $K'(T) = \xi$ has then no real root.

Conversely, suppose $K(T)$ exists for all T and $K'(T) = \xi$ has no real root when $\xi < a$ or $\xi > b$. Then $M'(T, \xi)$ has constant sign in the domains $\xi < a$, $-\infty < T < \infty$ and $\xi > b$, $-\infty < T < \infty$ so that $M(T, \xi)$ is monotonic in T for these values of ξ.

Moreover $M(T, \xi)$ must increase with T for all $\xi < a$, and decrease with T for all $\xi > b$. For if $M(T, \xi)$ increases with T, then $dF(x) = 0$ for every $x < \xi$, otherwise $M(-\infty, \xi) = \infty$ and if this were true for all $\xi > b$ we should have $dF(x) = 0$ for all x. Similarly $M(T, \xi)$ cannot decrease with T for $\xi < a$.

Hence when $\xi < a$, $dF(x) = 0$ for all $x < \xi$, that is $F(x) = 0$ for all $x < a$. In the same way $F(x) = 1$ for all $x > b$.

THEOREM 6.2. *Let $F(x) = 0$ for $x < a$, $0 < F(x) < 1$ for $a < x < b$, $F(x) = 1$ for $b < x$, where $-\infty < a < b < \infty$. Then for every ξ in $a < \xi < b$ there is a unique simple root T_0 of $K'(T) = \xi$. As T increases from $-\infty$ to ∞, $K'(T)$ increases continuously from $\xi = a$ to $\xi = b$.*

PROOF. When $a < \xi < b$, $M'(-\infty, \xi) = -\infty$ and $M'(\infty, \xi) = \infty$, and $M'(T, \xi)$ is strictly increasing with T since $M''(T) > 0$. So for each ξ in $a < \xi < b$ there is a unique root T_0 of $M'(T, \xi) = 0$ and hence of $K'(T) = \xi$. Also $K''(T_0) = M''(T_0, \xi)/M(T_0, \xi)$ so that $0 < K''(T_0) < \infty$, and T_0 is a simple root and $K'(T_0)$ is a strictly increasing function of T_0.

For all T, $M'(T, b) < 0$ and so $K'(T) < b$, but $M'(T, b - \epsilon) \to \infty$ as $T \to \infty$

for every $\epsilon > 0$ so that $K'(T) > b - \epsilon$ for all sufficiently large T. Hence $K'(T) \to b$ as $T \to \infty$. Similarly $K'(T) \to a$ as $T \to -\infty$. This also implies $K''(T) \to 0$ as $T \to \pm\infty$.

The theorem has an obvious interpretation in terms of the family of conjugate distributions (the term is due to Khinchin [14])

$$dF(x, T) = Ce^{Tx}dF(x)$$

which have mean $K'(T)$ and variance $K''(T)$.

A complication arises when a and b are allowed to be infinite. Suppose for example that a is finite but $b = \infty$, so that $K(T)$ exists in $-\infty < T < c_2$ where $0 \leqq c_2 \leqq \infty$. If $c_2 = \infty$, then $K'(T) \to \infty$ as $T \to \infty$ and the theorems still hold, for however large ξ is, $M'(T, \xi) \to \infty$ as $T \to \infty$ and so $K'(T) > \xi$ for all sufficiently large T.

But if $c_2 < \infty$ the corresponding theorems do not hold without a further condition, for it is not necessarily true that $K'(T) \to \infty$ as $T \to c_2$. Consider the class of distributions

$$dF(x) = e^{-c_2 x}\, dG(x)$$

where $\int_a^\infty dG(x) = m_0 < \infty$ and $\int_a^\infty x dG(x) = m_1 < \infty$, but $\int_a^\infty e^{\epsilon x}\, dG(x) = \infty$,

for all $\epsilon > 0$. Here $K'(T)$ increases from $-\infty$ to m_1/m_0 as T increases from $-\infty$ to c_2, but $K'(T) = \infty$ for all $T > c_2$. So for $\xi > m_1/m_0$, $K'(T) = \xi$ has no real root though the distribution may extend to ∞.

The case $a = -\infty$ can be discussed similarly. In the general case where $K(T)$ exists in $-c_1 < T < c_2$ and a and b may be infinite, the conditions

$$\lim_{T \to c_2} K'(T) = b, \qquad \lim_{T \to -c_1} K'(T) = a \tag{6.1}$$

are required for every ξ in (a, b) to have a corresponding T_0 in $(-c_1, c_2)$. They will be automatically satisfied except when a or b is infinite and the corresponding c_1 or c_2 is finite, in which case the appropriate condition has to be stated explicitly. But even when (6.1) is not satisfied the approximation $g_n(\bar{x})$ and the expansion (2.6) can still be used whenever $\bar{x}$ lies within the restricted range of values assumed by $K'(T)$.

7. Accuracy at the ends of the range of $\bar{x}$. We return to the distributions having a density function, and examine the accuracy of $g_n(\bar{x})$ and the expansion (2.6) for values of $\bar{x}$ near the ends of its admissible range (a, b), where the approximation might be expected to fail. It is assumed that the appropriate conditions hold for $K'(T) = \bar{x}$ to have a unique real root T_0 for every $\bar{x}$ in (a, b).

It has been proved that

$$| f_n(\bar{x})/g_n(\bar{x}) - 1 | < A(\bar{x})/n, \tag{7.1}$$

where $A(\bar{x})$ may depend on $\bar{x}$ since it is a function of T_0. The family of expansions (4.3) provides similar inequalities, and in particular an inequality of type (7.1) holds for symmetrical distributions when $g_n(\bar{x})$ is replaced by the limiting normal approximation to $f_n(\bar{x})$. But it is well known that the relative accuracy of the normal approximation, and of the Edgeworth series generally, deteriorates in most cases as $\bar{x}$ approaches the ends of its range. For example, if the interval (a, b) is finite and $f_n(\bar{x}) \to 0$ as $\bar{x} \to a$ or b, what corresponds to $A(\bar{x})$ in (7.1) becomes intolerably large as x approaches a or b, since the normal approximation can never be zero.

We now show that for a wide class of distributions $g_n(\bar{x})$ satisfies (7.1) with $A(\bar{x}) = A$, independent of $\bar{x}$, as $\bar{x}$ approaches a or b. In fact, for such distributions the asymptotic expansion of $f_n(\bar{x})/g_n(\bar{x})$ given by (2.6) is valid uniformly as $\bar{x} \to a$ or b. This will be so if $\lambda_j(T)$ remains bounded as $T \to -c_1$ or c_2 for every fixed j, so we examine the behaviour of $\lambda_j(T)$ near the ends of the interval. Equivalently, we study the conjugate distributions with density function

$$f(x, T) = Ce^{Tx}f(x) \tag{7.2}$$

whose jth cumulant is $K^{(j)}(T)$. The form of $f(x, T)$ as T approaches $-c_1$ or c_2 depends on the behaviour of $f(x)$ as x approaches a or b. For the commonest end conditions on $f(x)$, it will appear that $f(x, T)$ approximates either to the gamma form of Example 5.2 or to the normal form as $T \to -c_1$ or c_2. In the first case $\lambda_j(T)$ is bounded for given j; in the second case $\lambda_j(T) \to 0$ so that $g_n(\bar{x})$, for any n, becomes progressively more accurate as $\bar{x} \to b$, its relative error tending to a limiting value which is of smaller order than any power of n^{-1}.

We begin by discussing distributions with $b = \infty$ and first consider asymptotic forms of $f(x)$ when x is large for which $f(x, T)$ approximates to the gamma form.

EXAMPLE 7.1. $\qquad f(x) \sim Ax^{\alpha-1}e^{-cx}, \qquad \alpha > 0, c > 0.$

Let X be large. Then

$$M^{(j)}(T) = \int_{-\infty}^{\infty} x^j e^{Tx} f(x)\, dx = I_1 + I_2$$

where $I_1 = \int_{-\infty}^{X} x^j e^{Tx} f(x)\, dx$ is bounded as $T \to c$, and for small $c - T$,

$$I_2 \sim \int_X^{\infty} x^{j+\alpha-1} e^{-(c-T)x}\, dx = \frac{A}{(c-T)^{j+\alpha}} \int_{X(c-T)}^{\infty} w^{j+\alpha-1} e^{-w}\, dw$$

$$\sim A\,\Gamma(j+\alpha)/(c-T)^{j+\alpha}.$$

Thus

$$K^{(j)}(T) \sim \frac{\alpha}{(c-T)^j}; \qquad \lambda_j(T) \sim \alpha^{1-j/2} \tag{7.3}$$

for every j. In this case $f(x, T)$ tends to the gamma form as $T \to c$. The result is in fact a familiar Abelian theorem for Laplace transforms, and a more general form of it (Doetsch [7] p. 460) can be restated for our purpose as follows.

THEOREM 7.1. *Let $f(x) \sim Ax^{\alpha-1}l(x)e^{-cx}$ for $\alpha > 0$ and $c > 0$, where $l(x)$ is continuous and $l(kx)/l(x) \to 1$ as $x \to \infty$ for every $k > 0$. Then, as $T \to c$,*

$$M^{(j)}(T) \sim A\,\frac{\Gamma(j+\alpha)}{(c-T)^{j+\alpha}}\,l\left(\frac{1}{c-T}\right) \quad \textit{and} \quad \lambda_j(T) \sim \alpha^{1-j/2}.$$

This enables us to include end conditions of the form $Ax^{\alpha-1}\log x\, e^{-cx}$ or $Ax^{\alpha-1}\log\log x\, e^{-cx}$, etc. In all such cases $f(x, T)$ tends to the gamma form as $T \to c$.

In the second class of end conditions $f(x, T)$ approximates to the normal form for limiting values of T. We first consider heuristically some typical examples, again with $b = \infty$

EXAMPLE 7.2. $\quad f(x) \sim A \exp(\beta x^\alpha - cx), \qquad \beta > 0, c > 0, 0 < x < 1.$

Here we might expect $\lambda_j(T) \to 0$ as $T \to c$, for when $c - T$ is small the dominant part of $f(x, T)$ lies in the region of large x where

$$f(x, T) \sim CA \exp(\beta x^\alpha - (c - T)x).$$

This has a unique maximum at $x_0 = [\alpha\beta/(c - T)]^{1/(1-\alpha)}$ which is large for small $c - T$. If we put $x = x_0 y$ the corresponding density for y is $c' \exp[\beta x_0^\alpha (y^\alpha - \alpha y)]$ which has a sharp maximum at $y = 1$, near which it approximates to the normal form $c'' \exp[-\frac{1}{2}\beta\alpha(1-\alpha)x_0^\alpha(y-1)^2]$; it is relatively negligible elsewhere.

EXAMPLE 7.3. $\quad f(x) \sim A \exp(-\beta x^\alpha), \qquad \beta > 0, \alpha > 1.$

In this case T can be indefinitely large. We again expect $\lambda_j(T) \to 0$ as $T \to \infty$, for

$$f(x, T) \sim CA \exp[-\beta x^\alpha + Tx]$$

has a unique maximum at $x_0 = (T/\alpha\beta)^{1/(\alpha-1)}$ which tends to infinity with T; with $x = x_0 y$ the density for y becomes $c' \exp[\beta x_0^\alpha(y^\alpha - \alpha y)]$, which approximates to $c'' \exp[-\frac{1}{2}\beta\alpha(\alpha-1)x_0^\alpha(y-1)^2]$ as before.

These examples are included in the following general theorem concerning end conditions of the type $f(x) \sim e^{-h(x)}$, where $x^2h''(x) \to \infty$ as $x \to \infty$. Subject to a restriction on the variation of $h''(x)$ it is shown that $\lambda_j(T) \to 0$ in such cases as T tends to its upper limit.

THEOREM 7.2. *Let $f(x) \sim e^{-h(x)}$ for large x, where $h(x) > 0$ and $0 < h''(x) < \infty$. Let $v(x)$ and $w(x)$ exist such that*

$$\text{(i)} \quad [v(x)]^2h''(x) \to \infty \qquad \text{(ii)} \quad e^{-w(x)}h''(x) \to 0$$

monotonically as $x \to \infty$, where

$$v(x) > 0, \qquad |v'(x)| \leqq \alpha < \infty, \qquad w(x) = \int (1/v(x))\,dx.$$

Then $\lambda_j(T) \to 0$ as T tends to its upper limiting value.

Examples 7.2 and 7.3 are covered by $v(x) = x/\gamma$ for some $\gamma > 0$, conditions (i) and (ii) reducing to $x^2h''(x) \to \infty$, and $x^{-\gamma}h''(x) \to 0$. For $h(x) = e^x$ one can take $v(x) = \frac{1}{2}$, for $h(x) = e^{e^x}$ take $v(x) = \frac{1}{2}e^{-x}$, and so on. In all cases $v(x)/x$ is bounded and $w(x)$ increases at least as fast as $\log x$.

Since $0 < h''(x) < \infty$, $h'(x)$ is strictly increasing and $h'(x) \to c \leqq \infty$ as $x \to \infty$. Thus for large x, $f(x, T) \sim Ce^{Tx-h(x)}$ has a single maximum at the unique root x_0 of $h'(x_0) = T$, where $x_0 \to \infty$ as $T \to c \leqq \infty$.

The jth moment of $f(x, T)$ about x_0 is

$$\mu_j(T) = C \int_{-\infty}^{\infty} (x - x_0)^j f(x) e^{Tx}\, dx.$$

It will be shown that as $T \to c$ the major contribution to the integral comes from within a range $x_0 \pm \epsilon v(x_0)$ where ϵ is arbitrarily small. Consider first the behaviour of $v(x)$ and $w(x)$ in this interval. Since $|\, v'(x) \,| < \alpha$ as $x \to \infty$ we have for large x_0 and $|\, x - x_0 \,| \leqq \epsilon v(x_0)$,

$$|\, v(x) - v(x_0) \,| \leqq \alpha \,|\, x - x_0 \,| < \alpha\epsilon v(x_0)$$

that is

$$|\, v(x)/v(x_0) - 1 \,| < \alpha\epsilon. \tag{7.4}$$

Also for some x_1 in (x, x_0),

$$w(x) - w(x_0) = (x - x_0) w'(x_1) = (x - x_0)/v(x_1)$$

so that for $|\, x - x_0 \,| < \epsilon v(x_0)$,

$$|\, w(x) - w(x_0) \,| \leqq \epsilon \frac{v(x_0)}{v(x_1)} \leqq \frac{\epsilon}{1 - \alpha\epsilon}. \tag{7.5}$$

Let X be large, but choose T so that $x_0 > X + j/T$. Then

$$\mu_j(T) \sim C \int_{-\infty}^{X} (x - x_0)^j f(x) e^{Tx}\, dx$$
$$+ C \left\{ \int_X^{x_0 - \epsilon v(x_0)} + \int_{x_0-\epsilon v(x_0)}^{x_0+\epsilon v(x_0)} + \int_{x_0+\epsilon v(x_0)}^{\infty} \right\} [(x - x_0)^j \exp [xh'(x_0) - h(x)]\, dx$$
$$= I_1 + I_2 + I_3 + I_4$$

say. We examine the magnitude of each term as $T \to c$.

Since $(x_0 - x)^j e^{Tx}$ has its maximum at $(x_0 - j/T) > X$,

$$|\, I_1 \,| < C(x_0 - x)^j e^{TX} F(X) < C x_0^j e^{Xh'(x_0)}$$

For I_2,

$$|\, I_2 \,| = C e^{x_0 h'(x_0) - h(x_0)} \int_X^{x_0 - \epsilon v(x_0)} (x_0 - x)^j e^{-\psi(x, x_0)}\, dx \tag{7.6}$$

where we write $\psi(x, x_0) = h(x) - h(x_0) - (x - x_0)h'(x_0)$. By condition (ii) of

the theorem, when $x \leqq x_0$,

$$h''(x) \geqq e^{v(x)-v(x_0)}h''(x_0) > 0$$

and for $x \leqq x_0 - \epsilon v(x_0)$,

$$\begin{aligned} h'(x_0) - h'(x) &> h''(x_0) \int_{x_0-\epsilon v(x_0)}^{x_0} e^{v(x)-v(x_0)}\, dx \\ &\geqq \eta v(x_0)h''(x_0), \end{aligned}$$

where $\eta = \epsilon e^{-\epsilon/(1-\alpha\epsilon)}$, by (7.5). So $\psi(x, x_0) > \eta v(x_0)h''(x_0)(x_0 - x)$, and from (7.6),

$$| I_2 | < Ce^{x_0h'(x_0)-h(x_0)}/[\eta v(x_0)h''(x_0)]^{j+1}.$$

For I_3,

$$\begin{aligned} I_3 &= Ce^{x_0h'(x_0)-h(x_0)}\left\{\int_{x_0-\epsilon v(x_0)}^{x_0} + \int_{x_0}^{x_0+\epsilon v(x_0)}\right\} (x - x_0)^j e^{-\psi(x,x_0)}\, dx \\ &= Ce^{x_0h'(x_0)-h(x_0)}\{J_1 + J_2\}, \end{aligned}$$

say. When $x_0 - \epsilon v(x_0) \leqq x \leqq x_0$ we have from (i) and (ii),

$$e^{v(x)-v(x_0)} \leqq h''(x)/h''(x_0) \leqq [v(x_0)/v(x)]^2 \tag{7.7}$$

and so from (7.4) and (7.5),

$$\tfrac{1}{2}h''(x_0)(x - x_0)^2e^{-\epsilon/(1-\alpha\epsilon)} \leqq \psi(x, x_0) \leqq \tfrac{1}{2}h''(x_0)(x - x_0)^2(1 + \alpha\epsilon)^2$$

Putting $u = (x - x_0)[h''(x_0)]^{1/2}$ in J_1 makes the lower limit of integration become $-\epsilon v(x_0)[h''(x_0)]^{1/2}$, which tends to $-\infty$ as $x_0 \to \infty$ for fixed ϵ, by (i). Hence

$$J_1 \sim (-)^j \frac{2^{(j-1)/2}\Gamma[(j + 1)/2]}{[h''(x_0)]^{(j+1)/2}} \{1 + O(\epsilon)\}.$$

In the range $x_0 \leqq x \leqq x_0 + \epsilon v(x_0)$ the inequalities (7.7) are reversed and

$$\tfrac{1}{2}h''(x_0)(x - x_0)^2(1 - \alpha\epsilon)^2 \leqq \psi(x, x_0) \leqq \tfrac{1}{2}h''(x_0)(x - x_0)^2e^{\epsilon/(1-\alpha\epsilon)}$$

so that

$$J_2 \sim \frac{2^{(j-1)/2}\Gamma[(j + 1)/2]}{[h''(x_0)]^{(j+1)/2}} \{1 + O(\epsilon)\}.$$

Hence if j is even

$$I_3 \sim C \frac{e^{x_0h'(x_0)-h(x_0)}}{[h''(x_0)]^{(j+1)/2}} \frac{(2j)!}{2^j j!} (2\pi)^{1/2}\{1 + O(\epsilon)\}$$

while if j is odd

$$I_3 \sim C \frac{e^{x_0h'(x_0)-h(x_0)}}{[h''(x_0)]^{(j+1)/2}} \cdot O(\epsilon).$$

For I_4,

$$I_4 = e^{x_0 h'(x_0) - h(x_0)} \int_{x_0 + \epsilon v(x_0)}^{\infty} (x - x_0)^j e^{-\psi(x, x_0)} \, dx.$$

The inequality $h''(x) \geqq [v(x_0)/v(x)]^2 h''(x_0) > 0$ shows, as with I_2, that

$$I_4 < \frac{C e^{x_0 h'(x_0) - h(x_0)} j!}{[\epsilon(1 - \alpha\epsilon) v(x_0) h''(x_0)]^{j+1}}.$$

We now show that I_3 is the dominant term. First let j be even. As $T \to c$, both I_2/I_3 and I_4/I_3 are $O\{[v^2(x_0)h''(x_0)]^{-(j+1)/2}\}$ and so $\to 0$ for fixed ϵ. Further,

$$| I_1 |/I_3 < x_0^j [h''(x_0)]^{(j+1)/2} e^{h(X) - \psi(X, x_0)}.$$

From (ii), $h''(x_0) < e^{w(x_0)}$ as x_0 becomes large. Also since $v(x)/x$ is bounded, (i) implies that $(x - X)h''(x)v(x) \to \infty$ and so for all large enough x_0,

$$\psi(X, x_0) = \int_X^{x_0} (x - X)h''(x) \, dx > A \int_X^{x_0} (1/v(x)) \, dx = A\{w(x_0) - w(X)\}$$

whatever $A > 0$. Thus

$$| I_1 |/I_3 = O\{\exp [j \log x_0 - [A - \tfrac{1}{2}(j + 1)]w(x_0)]\}$$

which tends to zero as $T \to c$ for fixed X if A is large enough, since $w(x)$ increases at least as fast as $\log x$.

It follows that for even j,

$$\mu_j(T) \sim \frac{C e^{x_0 h'(x_0) - h(x_0)}}{[h''(x_0)]^{(j+1)/2}} \cdot \frac{(2j)!}{2^j j!} (2\pi)^{1/2}$$

$$\sim [h''(x_0)]^{-j/2} \frac{(2j)!}{2^j j!}$$

since $\mu_0(T) = 1$. Similarly when j is odd,

$$\mu_j(T) \sim [h''(x_0)]^{-j/2} \, O(\epsilon)$$

as $T \to c$, so the odd moments can be made relatively negligible for arbitrarily small ϵ. Thus the moments tend to those of the normal distribution and $\lambda_j(T) \to 0$ as $T \to c$.

Turning now to the case where $x \leqq b < \infty$ we consider forms of $f(x)$ when $b - x$ is small. Again there are found to be two classes of end conditions for which $\lambda_j(T)$ is bounded as $T \to \infty$, where $f(x, T)$ tends respectively to the gamma and to the normal form. It is convenient to put $u = b - x$ and regard $(-)^j K^{(j)}(T)$ as the jth cumulant of the distribution of u with density $f(b - u, T) = Be^{-Tu} f(b - u)$ for $u \geqq 0$.

EXAMPLE 7.4. $\quad f(x) \sim A(b - x)^{\alpha - 1}, \qquad \alpha > 0.$

The jth moment of u about its origin is

$$B \int_0^\infty u^j e^{-Tu} f(b-u)\, du \sim BA \int_0^\delta u^{j+\alpha-1} e^{-Tu}\, du + B \int_\delta^\infty u^j e^{-Tu} f(b-u)\, du$$

$$\sim BA \frac{\Gamma(\alpha+j)}{T^{\alpha+j}} \qquad T \to \infty,$$

where δ is small, the remainder being $O(e^{-T\delta})$ for large T. It follows that $\lambda_j(T) \sim (-)^j \alpha^{1-j/2}$. As in Example 7.1 this is a well known result on Laplace transforms, and its more general form (Doetsch [7], p. 476) yields the following theorem.

THEOREM 7.3. *Let $f(x) \sim A(b-x)^{\alpha-1} l(b-x)$ for $\alpha > 0$, where $l(u)$ is continuous and $[l(ku)]/l(u) \to 1$ as $u \to 0$ for every $k > 0$. Then $\lambda_j(T) \sim (-)^j \alpha^{1-j/2}$.*

For example $l(u)$ could be $\log(1/u)$ or $\log\log(1/u)$, etc.

The second class of end conditions is typified by the following example.

EXAMPLE 7.5. $\qquad f(x) \sim A \exp[-\beta/(b-x)^\gamma], \qquad \beta > 0, \gamma > 0.$

As in Example 7.2 we expect $\lambda_j(T) \to 0$ as $T \to \infty$, for

$$Ce^{-Tu} f(b-u) \sim CA \exp[-Tu - \beta/u^\gamma]$$

has a unique maximum at $u_0 = (\beta\gamma/T)^{1/(\gamma+1)}$, and the density function for $y = u/u_0$ is

$$C' \exp[-\beta u_0^{-\gamma}(\gamma y + y^{-\gamma})] \sim C'' \exp[-\tfrac{1}{2}\beta\gamma(\gamma+1)u_0^{-\gamma}(y-1)^2]$$

The general theorem analogous to Theorem 7.2 is:

THEOREM 7.4. *Let $f(x) \sim e^{-h(x)}$ for small $b-x$, where $h(x) > 0$ and $0 < h''(x) < \infty$. Let $v(u)$ and $w(u)$ exist such that*

$$\text{(i) } [v(b-x)]^2 h''(x) \to \infty, \qquad \text{(ii) } e^{w(b-x)} h''(x) \to 0,$$

monotonically as $x \to b$, where $v(0) = 0$ and $w(u) = \int [1/v(u)]\, du$, and $0 < v'(u) \leq \alpha < \infty$ for $u > 0$. Then $\lambda_j(T) \to 0$ as $T \to \infty$.

As before $h'(x)$ is strictly increasing, and $h'(x) \to \infty$ as $x \to b$ since (i) implies $(b-x)^2 h''(x) \to \infty$, and $h'(x_0) = T$ has a unique root x_0 where $x_0 \to b$ as $T \to \infty$. Thus $f(b-u, T)$ has a unique maximum at $u_0 = b - x_0$ for large T, and $u_0 \to 0$ as $T \to \infty$. The jth moment of u about u_0 is

$$\mu_j(T) = B \int_0^\infty (u-u_0)^j e^{-Tu} f(b-u)\, du.$$

We write

$$\int_0^\infty = \int_0^{u_0 - \epsilon v(u_0)} + \int_{u_0-\epsilon v(u_0)}^{u_0+\epsilon v(u_0)} + \int_{u_0+\epsilon v(u_0)}^{\delta} + \int_\delta^\infty$$

where ϵ and δ are small. The proof then proceeds with appropriate modifications as in Theorem 7.2.

8. Discrete variables. The discussion has so far been concerned with approximations to probability densities, but the saddlepoint method provides similar approximations to probabilities when the variable is discrete, and indeed it is typically used for this purpose in statistical mechanics. Consider, for example, a variable x which takes only integral values $x = r$ with nonzero probabilities $p(r)$. The moment generating function,

$$M(T) = e^{K(T)} = \sum_r p(r)e^{Tr} \tag{8.1}$$

is assumed to satisfy conditions (6.1).

The mean $\bar{x}$ of n independent x's can take only values $\bar{x} = r/n$, for which the probabilities are

$$p_n(\bar{x}) = \frac{1}{2\pi i}\int_{\tau-i\pi}^{\tau'+i\pi} e^{n[K(T)-T\bar{x}]}\,dT \tag{8.2}$$

analogous to (2.2). The contour is again chosen to be the line $T = T_0 + iy$ passing through the unique real saddle point T_0, but it now terminates at $T_0 \pm i\pi$. This ensures that the integrand attains its maximum modulus at T_0 but nowhere else on the contour, provided we exclude cases where $p(r) = 0$ except at multiples of an integer greater than unity. The discussion of Section 2 shows that the maximum modulus is attained when y satisfies $\cos(ry - \alpha) = 1$ for some α and all integral r, and $y = 0$ is the only possible value in $(-\pi, \pi)$. The argument then proceeds as before and leads to the approximation

$$p_n(\bar{x}) \sim \frac{e^{n[K(T_0)-T_0\bar{x}]}}{[2\pi nK''(T_0)]^{1/2}}\{1 + O(n^{-1})\} \tag{8.3}$$

where $\bar{x} = r/n$ and r is an integer.

As an example, consider the binomial distribution

$$p(r) = \binom{N}{r}(1 - p)^r p^{n-r}.$$

Here

$$K(T) = N\log\{1 + p(e^T - 1)\}, \qquad K'(T_0) = Npe^{T_0}/[1 + p(e^{T_0} - 1)] = \bar{x},$$

$$e^{T_0} = [\bar{x}/(N - \bar{x})]\cdot[(1 - p)/p], \quad K''(T_0) = \bar{x}(N - \bar{x})/N,$$

$$p_n(\bar{x}) \sim \frac{N^{nN+1/2}}{(2\pi n)^{1/2}}\,\frac{(1 - p)^{n(N-\bar{x})}p^{n\bar{x}}}{(N - \bar{x})^{n(N-\bar{x})+1/2}\,\bar{x}^{n\bar{x}+1/2}}\{1 + O(n^{-1})\}.$$

This is the familiar intermediate form obtained on replacing the factorials by Stirling's approximation before passing to the normal limit.

9. Ratio of sums of random variables. The saddlepoint technique can also be applied to the distribution of ratios. Cramér (4) has shown that if x and y are two independent random variables with densities $f_1(x)$ and $f_2(y)$ and characteristic functions $\phi_1(t)$ and $\phi_2(u)$, and if $y \geq 0$, the density function for $r = x/y$ is

given by

$$f(r) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \phi_1(t)\phi_2'(-rt)\, dt \tag{9.1}$$

provided y has a finite mean. (Gurland [11] relaxes this condition by introducing principal values. Cramér states the condition differently and appears to require unnecessarily that x shall have a finite mean also.) Cramér deduced the result from the distribution of $x - ry$ for fixed r, but it also follows on applying Parseval's theorem to the formula

$$f(r) = \int_0^{\infty} f_1(ry)f_2(y)y\, dy \tag{9.2}$$

where y must have a finite mean to make $\phi_2'(-rt)$ the Fourier transform of $iyf_2(y)$. In terms of cumulant generating functions (9.1) takes the form

$$f(r) = \frac{1}{2\pi i} \int_{\tau-i\infty}^{\tau'+i\infty} e^{K_1(T)+K_2(-rT)}\, K_2'(-rT)\, dT.$$

Let $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ be independent random samples from these distributions, their sums being X and Y. The density for $R = x/y$ is then

$$f_{n_1,n_2}(R) = \frac{n_2}{2\pi i} \int_{\tau-i\infty}^{\tau'+i\infty} e^{n_1K_1(T)+n_2K_2(-RT)}\, K_2'(-RT)\, dT.$$

When n_1 and n_2 are large, an approximation is found by passing the path of integration through a saddlepoint T_0 of the exponential part of the integrand, given by

$$n_1K_1'(T_0) - n_2RK_2'(-RT) = 0 \tag{9.3}$$

Assuming conditions (6.1) to be satisfied, both $K_1'(T)$ and $K_2'(T)$ are increasing functions of T taking every admissible value of X and Y respectively as T varies over its appropriate interval, so that to every R there is a single real root T_0 of (9.3). (However, it is possible for the same T_0 to correspond to more than one value of R, since $TK_2'(T)$ is not necessarily monotonic and so dT_0/dR may change sign). Proceeding as before, expanding $K_2'(-RT)$ also, we obtain an asymptotic expansion whose dominant term is

$$g_{n_1,n_2}(R) = \frac{n_2K_2'(-RT_0)e^{n_1K_1(T_0)+n_2K_2(-RT_0)}}{\{2\pi[n_1K_1''(T_0) + n_2R^2K_2''(-RT_0)]\}^{1/2}}$$

the remainder being relatively $O(n^{-1})$ where $n = \min(n_1, n_2)$.

EXAMPLE 9.1. $f_1(x) = A_1x^{\alpha_1-1}e^{-\beta_1x}, \quad f_2(y) = A_2y^{\alpha_2-1}e^{-\beta_2y}$,

where x, y, α_1, β_1, α_2, β_2 are all positive. In this case

$$T_0 = \frac{1}{R}\frac{(n_2\alpha_2\beta_1R - n_1\alpha_1\beta_2)}{(n_1\alpha_1 + n_2\alpha_2)}.$$

The approximation is found to be

$$g_{n_1,n_2}(R) = \frac{\beta_1^{n_1\alpha_1}\beta_2^{n_2\alpha_2}(n_1\alpha_1 + n_2\alpha_2)^{n_1\alpha_1+n_2\alpha_2-1/2}}{(2\pi)^{1/2}(n_1\alpha_1)^{n_1\alpha_1-1/2}(n_2\alpha_2)^{n_2\alpha_2-1/2}} \frac{R^{n_1\alpha_1-1}}{(\beta_1 R + \beta_2)^{n_1\alpha_1+n_2\alpha_2}},$$

which differs from the exact density function only in the normalising constant, and so is "exact" in the sense of Example 5.2. This suggests that there may again be a class of distributions for which the relative error is bounded uniformly over the whole range of R for every n.

An extension of (9.1) is available when the variables are not independent (Cramér [6] p. 317, ex. 6; Geary [10]). If (x, y) has a bivariate density function $f(x, y)$ everywhere and characteristic function $\phi(t, u)$, and if $y \geqq 0$, the density function for $r = x/y$ is

$$f(r) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\left[\frac{\partial\phi(t, u)}{\partial u}\right]_{u=-rt} dt \tag{9.4}$$

provided the integral is absolutely convergent, which requires y to have a finite mean. The following proof of (9.4) shows the integrand to be proportional to a characteristic function which attains its maximum modulus only at $t = 0$, so that the previous methods are applicable. Corresponding to (9.2) we have

$$f(r) = \int_0^{\infty} f(ry, y)y\, dy. \tag{9.5}$$

Write $\eta = E(y)$ and define a new distribution with density and characteristic function

$$h(x, y) = \frac{1}{\eta} yf(x, y) \qquad \phi(x, y) = \frac{1}{\eta}\frac{\partial\phi(t, u)}{i\partial u}. \tag{9.6}$$

From (9.5) it is seen that $(1/\eta)f(r)$ can be regarded as the probability density at zero of the variable $w = x - ry$, where (x, y) has the distribution (9.6). The result then follows from the fact that w has the characteristic function

$$\frac{1}{\eta}\left[\frac{\partial\phi(t, u)}{i\partial u}\right]_{u=-rt}.$$

For a random sample of n, the ratio R of the sums X and Y has density

$$f_n(R) = \frac{n}{2\pi i}\int_{\tau-i\infty}^{\tau'+i\infty} e^{nK(T,-RT)}\left[-\frac{1}{T}\frac{\partial K(T, -RT)}{\partial R}\right]\partial T$$

in terms of the bivariate cumulant generating function. The saddlepoint approximation is

$$g_n(R) = \left\{\frac{n}{2\pi K''(T_0, -RT_0)}\right\}^{1/2} e^{nK(T_0,-RT_0)}\left[\frac{-1}{T_0}\frac{\partial K(T_0, -RT_0)}{\partial R}\right]$$

where

$$\frac{\partial K(T_0, -RT_0)}{\partial T_0} = 0.$$

EXAMPLE 9.2. Let $x = \frac{1}{2}u^2$ and $y = \frac{1}{2}v^2$, where u and v have a bivariate normal distribution with unit variances and correlation coefficient ρ. Thus $R = X/Y$ is a "variance ratio" calculated from two equal correlated samples. The exact distribution of R has been given by Bose [1] and Finney [8]. We find

$$K(T, -RT) = \tfrac{1}{2} \log \{1 + (R - 1)T - RT^2(1 - \rho^2)\},$$

$$T_0 = \frac{(R-1)}{2R(1-\rho^2)}, \qquad K''(T_0, -RT_0) = \frac{4R(1-\rho^2)^2}{[(1+R)^2 - 4\rho^2 R]},$$

$$\frac{-1}{T_0} \frac{\partial K(T_0, -RT_0)}{\partial R} = \frac{(R+1)(1-\rho^2)}{[(1+R)^2 - 4\rho^2 R]},$$

$$g_n(R) = 2^{n-1} \left(\frac{n}{2\pi}\right)^{1/2} \frac{(1-\rho^2)^{n/2} R^{(n/2)-1}(1+R)}{[(1+R)^2 - 4\rho^2 R]^{(n+1)/2}}$$

which again agrees with the exact distribution except for the normalising constant.

In the most general situation where the sample members are themselves correlated, the saddlepoint method can still be applied. In each particular case the contribution to the integral from parts of the contour outside a neighbourhood of the saddlepoint must be established as negligible. One can obtain, in this way an approximation to the distribution of the sample serial correlation coefficient of lag 1 from a linear Markoff population. With the usual "circular" definitions it turns out to be the approximation given by Leipnik [15], but a similar approximation can also be found for the noncircular case. A detailed account of this work will appear elsewhere.

10. Acknowledgments. I am much indebted to D. V. Lindley and Sir Harold Jeffreys for many stimulating discussions and useful suggestions, to a referee for comments which led to improvements in the paper, and to D. A. East for computing Table 1.

11. Appendix. The identity of the series (2.6) and 3.3) may be established as follows. For the contour $T = T_0 + iy$ the inversion formula is

$$f_n(\bar{x}) = \frac{n}{2\pi} e^{n[K(T_0) - T_0 \bar{x}]} \int_{-\infty}^{\infty} e^{-nw^2/2}\, dy$$

where w^2 is defined by (3.2). With $v = y[nK''(T_0)]^{1/2}$ and $s = n^{-1/2}$ this becomes

$$f_n(\bar{x}) = \frac{1}{2\pi} \left[\frac{n}{K''(T_0)}\right]^{1/2} e^{n[K(T_0) - T_0 \bar{x}]} \int_{-\infty}^{\infty} e^{-w^2(ivs)/2s^2}\, dv$$

with $z = ivs$ in (3.2). To get (2.5) the integrand is expanded as a power series in s. Term-by-term integration gives (2.6). Thus

$$\exp\left[-\frac{w^2(ivs)}{2s^2}\right] = \sum_{m=0}^{\infty} b_m(v) s^m$$

where

$$b_m(v) = \frac{1}{m!}\frac{\partial^m}{\partial s^m}\exp\left[-\frac{w^2(ivs)}{2s^2}\right]\bigg|_{s=0}$$
$$= \frac{1}{m!}v^m\frac{\partial^m}{\partial x^m}\exp\left[-\frac{v^2w^2(ix)}{2x^2}\right]\bigg|_{x=0}$$

Since $w^2(ix)/x^2 \sim 1 + O(x)$, for small x we can interchange the order of differentiation with respect to x and integration with respect to v. Only the even terms survive and

$$\int_{-\infty}^{\infty} b_{2r}(v)\,dv = \frac{1}{(2r)!}\frac{d^{2r}}{dx^{2r}}\int_{-\infty}^{\infty} v^{2r}\exp\left[-\frac{v^2w^2(ix)}{2x^2}\right]dv\bigg|_{x=0}$$
$$= \frac{(2\pi)^{1/2}}{2^r r!}\frac{d^{2r}}{dx^{2r}}\left[\frac{x}{w(ix)}\right]^{2r+1}\bigg|_{x=0} = (2\pi)^{1/2}a_r,$$

putting $z = ix$ in (3.4).

REFERENCES

[1] S. Bose, "On the distribution of the ratio of variance of two samples drawn from a given normal bivariate correlated population," *Sankhya*, Vol. 2 (1935), pp. 65–72.

[2] D. R. Cox, "A note on the asymptotic distribution of the range," *Biometrika*, Vol. 35 (1948), pp. 311–315.

[3] H. Cramér, "On the composition of elementary errors," *Skand. Aktuarietids.*, Vol. 11 (1928), pp. 13–74, 141–180.

[4] H. Cramér, *Random Variables and Probability Distributions*, Cambridge University Press, 1937.

[5] H. Cramér, "Sur un nouveau théorème-limite de la théorie des probabilités," *Actualités Scientifiques et Industrielles*, No. 736, Hermann & Cie, Paris, 1938.

[6] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[7] G. Doetsch, *Handbuch der Laplace Transformation*, Band I, Birkhäuser, Basel, 1950.

[8] D. J. Finney, "The distribution of the ratio of estimates of two variances in a sample from a normal bivariate population," *Biometrika*, Vol. 30 (1938), pp. 190–192.

[9] R. H. Fowler, *Statistical Mechanics*, Cambridge University Press, 1936.

[10] R. C. Geary, "Extension of a theorem by Harald Cramér on the frequency distribution of a quotient of two variables," *J. Roy. Stat. Soc.*, Vol. 17 (1944), pp. 56–57.

[11] J. Gurland, "Inversion formulae for the distribution of ratios," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 228–237.

[12] H. Jeffreys, *Theory of Probability*, Oxford University Press, 1948.

[13] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, Cambridge University Press, 1950.

[14] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover Publications, Inc., New York, 1949.

[15] R. B. Leipnik, "Distribution of the serial correlation in a circularly correlated universe," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 80–87.

[16] H. L. Seal, "Spot the prior reference," *J. Inst. of Actuaries, Students' Soc.*, Vol. 10 (1951), pp. 255–258.

[17] G. N. Watson, *Theory of Bessel Functions*, Cambridge University Press, 1948.

[18] A. Wintner, *The Fourier Transforms of Probability Distributions*, Johns Hopkins University, Baltimore, 1947.

A GENERAL THEORY OF DISCRIMINATION WHEN THE INFORMATION ABOUT ALTERNATIVE POPULATION DISTRIBUTIONS IS BASED ON SAMPLES

By C. Radhakrishna Rao

Indian Statistical Institute

1. Introduction. The problem of discrimination, that is of assigning an observed individual to its proper group, admits a simple solution when the distributions of measurements in the alternative populations are completely specified. Research in this direction originated with the use of the linear discriminant function introduced in 1936 by Fisher [3]. In 1939 Welch [24] showed that a general discriminant function in the case of two alternatives is the likelihood ratio of the two hypotheses, and is deducible either from Bayes' theorem with given a priori probabilities or by the use of a lemma by Neyman and Pearson [11] when the errors for the two hypotheses are minimised in any given ratio.

A general theory of decision functions when the alternatives are finite or infinite was developed by Wald [19] in 1939 and further generalized by him in 1949 [23]. In 1945 von Mises [9] obtained, in the case of a finite number of alternatives, the solution to the problem of minimising the maximum error, which is the general theme of Wald's work. Explicit solutions of Bayes' form, with given a priori probabilities or ratio of errors for the alternative groups, and the construction and use of a doubtful region were discussed by the author [13] in 1948. Related problems and the extension to problems of selection have been treated in a subsequent series of papers [15], [16].

In all these cases the alternative population distributions are assumed to be completely specified. The decision rule consists in setting up a correspondence between values observed in a sample and the alternative population distributions. In practice it is rarely possible to specify completely the distributions, but they may be estimable on the basis of independent samples from each of the alternative distributions.

Let $S_1, \cdots, S_k$ be independent samples from k alternative populations which may be partially specified, as when the functional forms of the probability densities are given but with unspecified parameters, or completely unspecified. After a sample S is drawn from a population known a priori to be one of the above set of k populations, the problem is to infer from which population the sample S has been drawn. The decision rule should be in the form of associating S with one of the samples $S_1, \cdots, S_k$, and declaring that S has come from the same population as the sample with which it is associated.

The usual practice is to estimate the alternative distributions on the basis of the sample information, and to use them in the solution which is strictly applicable when the alternatives are completely specified. This is probably the right

Received 8/6/53, revised 1/22/54.

approach when estimation is based on large samples. Fix and Hodges [4] have shown that this procedure is consistent under certain conditions, that is, with probability tending to unity it gives the same results as when the alternatives are known, provided the sample sizes are indefinitely increased. This procedure can be shown to be asymtotically the best in the sense of Wald [20].

No systematic attempt seems to have been made to offer solutions for finite samples. Wald [22] proposed to solve this problem in the case of two alternatives by obtaining the distribution of the estimated likelihood ratio or the linear discriminant function of Fisher. Even if the distribution problem is satisfactorily solved, it cannot be applied in practice since it involves unknown parameters.

In this paper some general methods have been developed with the help of which the discrimination problem can be solved, utilizing only the sample information. This theory is immediately applicable when the alternative distributions have given functional forms but with unspecified parameters. The nonparametric cases can be treated in a similar manner, but no attempt has been made in this paper to offer explicit solutions.

2. Statement of the problem. Let $p_1(x \mid \theta_1), \cdots, p_k(x \mid \theta_k)$ be k probability densities with known functional forms but unknown parameters. In the representation of the function $p(x \mid \theta)$, x stands for all the measurements and θ for all the unknown parameters. We have, in general, to deal with p-variate populations so that x stands for a vector of p stochastic variables. Samples of sizes $n_1, \cdots, n_k$ are available from these k populations. The observations from the ith population, for $i = 1, \cdots, k$, are denoted by

$$S_i: \quad x_j^i = (x_{1j}^i, \cdots, x_{pj}^i), \qquad j = 1, \cdots, n_i. \tag{2.1}$$

An individual known a priori to belong to one of the k groups has the measurements

$$S: \quad x = (x_1, \cdots, x_p). \tag{2.2}$$

The problem is to assign this individual to its proper group on the basis of the information supplied only by the observations (2.1) and (2.2), without making any assumption about the unknown parameters. The problem is similar when, instead of p measurements on a single individual, the sample S in (2.2) consists of p measurements on each of n individuals drawn from that population. The problem is to decide on the population from which S has arisen, using the information supplied by S and $S_1, \cdots, S_k$ of (2.1).

3. Some observations on the solution when the parameters are known. If the a priori probabilities of the observed individual belonging to the k groups are $\pi_1, \cdots, \pi_k$, then the Bayes' solution which minimises the errors of wrong classification is to assign individuals with measurements x to the ith group if $\pi_i p_i(x \mid \theta_i)$ has the highest value in the set

$$\pi_1 p_1(x \mid \theta_1), \cdots, \pi_k p_k(x \mid \theta_k). \tag{3.1}$$

The solution which assigns the individual to the ith group if $a_i p_i(x \mid \theta_i)$ has the highest value in the set

$$a_1 p_1(x \mid \theta_1), \cdots, a_k p_k(x \mid \theta_k) \tag{3.2}$$

has the property of minimising the frequencies of wrong classification for the various groups in a ratio determined by the procedure (3.2). This ratio is a function of $a_1, \cdots, a_k$; if possible, the constants may be chosen for any specified ratio [13].

When $\pi_1, \cdots, \pi_k$ are unknown or when the consideration of a priori probabilities is irrelevant, we have to depend on solution (3.2). One method is to choose the constants such that the errors are in an equal ratio, using the criterion of minimax [9], [23]. Another method is to choose $a_i = 1$ for all i, using the principle of maximum likelihood. The latter method gives an unbiased division of the space, that is, the probability with respect to the density p_j of all observation points assigned to the ith population is the highest for $j = i$. All Bayes' solutions do not have this property except in the case of two alternative populations. Also, it is not evident whether the minimax solution is always unbiased in the above sense. Some criterion has to be developed for the choice of a rule from the subclass of Bayes' solutions which are unbiased.

4. Large sample theory. The observations (2.1) and (2.2) considered in Section 2 can be represented by a point in a space of $(n_1 + n_2 + \cdots + n_k + 1)p$ or more generally of $(n_1 + n_2 + \cdots + n_k + n)p$ dimensions. Every division of the space into k regions $R_1, \cdots, R_k$ provides a decision rule, by which the ith population is accepted when the points fall in the corresponding region R_i.

The probability of correct classification β_i' for the ith group is the density of the region R_i when the last observation (2.2) arises from the ith group. If the a priori probability that the last observation belongs to the ith group is π_i, then the probability of correct classification is

$$\pi_1 \beta_1' + \cdots + \pi_k \beta_k'. \tag{4.1}$$

This is obviously less than $\pi_1\beta_1 + \cdots + \pi_k\beta_k$, where β_i are the values associated with the solution (3.1) when all the parameters are known a priori and samples do not provide any additional information.

Expression (4.1) is a function of the unknown parameters $(\theta_1, \cdots, \theta_k)$ and of the division $\mathfrak{D}$ of the space of $N = (n_1 + n_2 + \cdots + n_k + 1)p$ dimensions. This function is denoted by $f_N(\mathfrak{D}, \theta_1, \cdots, \theta_k)$ or simply by $f_N(\mathfrak{D}, \theta)$. Let $L_N(\mathfrak{D}_1, \mathfrak{D}_2)$ represent the least upper bound of the difference $f_N(\mathfrak{D}_1, \theta) - f_N(\mathfrak{D}_2, \theta)$ corresponding to two divisions $\mathfrak{D}_1$ and $\mathfrak{D}_2$.

Following a concept due to Wald [20] we define a sequence of divisions $\mathfrak{D}^*$ to be *asymptotically best* if there does not exist any other sequence $\mathfrak{D}$ such that

$$\limsup_{n_i \to \infty} L_N(\mathfrak{D}, \mathfrak{D}^*) > 0. \tag{4.2}$$

If there exists a sequence of divisions $\mathfrak{D}_u$ such that

$$(4.3) \qquad f_N(\mathfrak{D}_u, \theta) \to \pi_1\beta_1 + \cdots + \pi_k\beta_k \qquad \text{as } n_i \to \infty$$

uniformly in the parameters as the sample sizes individually tend to infinity, then such a sequence automatically satisfies the criterion (4.2) for being best asymptotically. Fix and Hodges [4] have shown that for the solution

$$(4.4) \qquad R_i: \quad \pi_i p_i(x \mid \hat{\theta}_i) \geqq \pi_j p_j(x \mid \hat{\theta}_j), \qquad j = 1, \cdots, k,$$

where $\hat{\theta}_1, \cdots, \hat{\theta}_k$ are uniformly consistent estimates of the parameters the probability of correct classifications, tends uniformly to $\pi_1\beta_1 + \cdots + \pi_k\beta_k$ as each sample size tends to infinity, provided the probability densities satisfy some mild regularity conditions. This result, together with property of uniform consistency of maximum likelihood estimates (true under some general conditions stated by Wald, [21], provides a method of constructing an asymptotically best solution of the type (4.4).

5. Small sample theory. Let us first consider the problem of two alternative groups. There are n_1 observations from the first group, n_2 from the second, and a single observation (each observation means a set of p measurements) on an individual whose group is unknown. If θ_1 and θ_2 are the parameters for the first and second groups, then the parameters applicable to the three sets of observations are

$$H_1: \qquad (\theta_1, \theta_2, \theta_1)$$

when the individual belongs to the first group and

$$H_2: \qquad (\theta_1, \theta_2, \theta_2)$$

otherwise. The two alternative hypotheses from which one is to be chosen on the basis of observations are, therefore, the vectors $(\theta_1, \theta_2, \theta_1)$ and $(\theta_1, \theta_2, \theta_2)$, whatever θ_1 and θ_2 may be.

5.1. *Test for H_2 against H_1 at a fixed significance level.* Let us choose one of these hypotheses (say H_2) as null and test it against the alternative H_1. For this we need critical regions in the space of $(n_1 + n_2 + 1)p$ observations which are similar with respect to the parameters θ_1 and θ_2 under the hypothesis $(\theta_1, \theta_2, \theta_2)$. Out of these, one which maximises the power with respect to the alternatives $(\theta_1, \theta_2, \theta_1)$ is to be chosen. How far this method yields successful results may be judged by a simple example.

Let $p_1(x \mid \theta_1)$ and $p_2(x \mid \theta_2)$ be univariate normal probability densities with unknown mean values θ_1 and θ_2 and unit standard deviation. From each population n observations are taken; the mean values are found to be $\bar{x}_1$ and $\bar{x}_2$. According to the null hypothesis, the last observation x belongs to the second group. In this case

$$T_1 = \bar{x}_1, \qquad T_2 = (x + n\bar{x}_2) / (1 + n)$$

are sufficient for θ_1 and θ_2. The critical region similar with respect to θ_1 and θ_2 has a conditional size α on the surfaces of constant values of T_1 and T_2.

If to these statistics is added $T_3 = x - \bar{x}_2$, then it is necessary to consider only the conditional distribution of T_3 given T_1, T_2. In fact, T_3 is distributed independently of T_1, T_2 under both hypotheses and has the densities proportional to

$$\exp\left\{\frac{-n}{2(n+1)}\,(T_3 - \overline{\theta_1 - \theta_2})^2\right\}, \qquad \exp\left\{\frac{-n}{2(n+1)}\,T_3^2\right\},$$

whose ratio is independent of the observed values from the first group.

The test derived above is the same as that for testing whether the observation x comes from the second group when the alternatives are unspecified. The situation is somewhat unfortunate in that the test does not utilize the information given by the second sample. Perhaps it is inevitable, if we have to come to decisions independently of any a priori knowledge restricting to a fixed significance level. This, however, suggests an intuitive approach to the problem of classification.

Suppose that it is possible to test the hypothesis that the individual belongs to a specified group, say the ith, (ignoring the fact that the alternatives are confined to a finite number about which we have some information) at any given probability level of rejection, and that all the critical regions corresponding to different probability levels are well ordered, the bigger containing the smaller. We define by ξ_i the least probability level at which the ith hypothesis can be rejected. The k groups supply k values $\xi_1, \cdots, \xi_k$, and it appears to be a reasonable rule to assign the individual to the jth group if ξ_j is the maximum in the set. The optimum properties of this rule will naturally depend on the nature of tests of the above hypotheses, but this is generally applicable in situations where reasonable tests exist.

Consider for example the univariate case where the k samples provide the averages $\bar{x}_1, \cdots, \bar{x}_k$ based on sizes $n_1, \cdots, n_k$ and pooled variance s^2 based on $(\sum n_i - k)$ degrees of freedom. If x is the observation on an individual to be classified, we calculate the probabilities

$$\xi_i = P\{\,|t| > |x - \bar{x}_i| \,/\, s\sqrt{1 + 1/n_i}\},$$

where the variable t has Student's distribution based on $(\sum n_i - k)$ degrees of freedom. The individual is assigned to that group for which ξ_i is a maximum. This rule is immediately applicable since it involves no new technique. Only a reasonable test should exist and the probability integral table should be available. It is, however, not easy to say what optimum properties are implied by this rule, except that errors are less for groups with larger sample sizes.

Another intuitive method which may yield fruitful results is to use fiducial probability distributions if they exist (as defined by Fisher [2]) of the observation x. Corresponding to k groups we can set up the k alternative fiducial distributions, using the samples. These distributions are parameter-free and the problem now

reduces to the classical case of assigning the observation x to one of k populations whose distributions are completely defined. It would, however, be somewhat difficult to study the optimum properties of this procedure.

In the following we will lay down a few postulates concerning the nature of the decision rule, and obtain solutions which have optimum properties when the alternative hypotheses are close to one another.

5.2. *A general postulate concerning the decision rule.* Let us denote the probability density of the observations from the ith group by

$$P_i(x^i \mid \theta_i) = p_i(x_1^i \mid \theta_i) \cdots p_i(x_{n_i}^i \mid \theta_i), \qquad i = 1, \cdots, k.$$

For simplicity we shall consider only nonrandomised decision rules which need a division of the sample space of $(n_1 + \cdots + n_k + 1)p$ dimensions into mutually exclusive regions $R_1, \cdots, R_k$. The rule of behaviour is to accept the hypothesis that the individual belongs to the ith population when the sample point falls in R_i.

In developing the arguments we shall choose the case of two alternative populations only, the conclusions being the same for several. In this problem there are two regions R_1 and R_2. The proportion of errors committed when the individual belongs to the first group is

$$\alpha_1(\theta_1, \theta_2) = \int_{R_2} P_1(x^1 \mid \theta_1) \quad P_2(x^2 \mid \theta_2) \quad p_1(x \mid \theta_1)\, dv.$$

Similarly for the other group,

$$\alpha_2(\theta_1, \theta_2) = \int_{R_1} P_1(x^1 \mid \theta_1) \quad P_2(x^2 \mid \theta_2) \quad p_2(x \mid \theta_2)\, dv.$$

Suppose that we need a decision rule for which the linear compound of errors

$$\pi_1\alpha_1(\theta_1, \theta_2) + \pi_2\alpha_2(\theta_1, \theta_2) \tag{5.2.1}$$

is a minimum. The compounding coefficients π_1 and π_2 may be assigned a priori probabilities, or suitable weights may be attached to the errors. If there exists a division of the space which minimises (5.2.1) irrespective of the true values of the parameters, then such a division cannot be improved upon. The minimum value of (5.2.1) for any given values θ_1 and θ_2 of the parameters is attained for the regions

$$R_1\colon\ \pi_1 p_1(x \mid \theta_1) \geqq \pi_2 p_2(x \mid \theta_2), \qquad R_2\colon\ \pi_2 p_2(x \mid \theta_2) \geqq \pi_1 p_1(x \mid \theta_1).$$

If the boundary of these regions is independent of the parameters θ_1 and θ_2, then we have a uniformly best division of the space. In such a case the sample observations do not provide any additional information.

If we exclude such special cases, it would appear that whatever may be the set of regions offered it will not be uniformly the best for all values of the unknown parameters and can be good only in some restricted sense. We need then some reasonable postulates governing the choice of a decision rule.

An obvious requirement on the decision rule is that it should not lead to contradictions or give recognizably bad results in particular cases. Let us consider the degenerate case when the two alternative distributions are identical, that is, $\theta_1 = \theta_2 = \theta$. For any division R_1, R_2 of the space, the errors committed for the two groups in this situation are $\alpha_1(\theta, \theta)$ and $\alpha_2(\theta, \theta)$, with the necessary condition $\alpha_1(\theta, \theta) + \alpha_2(\theta, \theta) = 1$. When the population distributions are equal the only rule is to assign individuals at random, subject to a given or a chosen frequency of errors for the two groups. It seems therefore reasonable to postulate that $\alpha_1(\theta, \theta)$ and $\alpha_2(\theta, \theta)$ should be constant independently of the common values of the parameters θ_1 and θ_2.

Further, let us imagine that for a given division of the space the value of $\alpha_1(\theta, \theta)$ at a neighbouring value $(\theta + \delta\theta)$ is more, implying that

$$\delta\theta \left\{ \frac{\partial}{\partial\theta_1} \alpha_1(\theta_1, \theta_2) + \frac{\partial}{\partial\theta_2} \alpha_1(\theta_1, \theta_2) \right\}_{\theta_1=\theta_2=\theta} = \delta\theta(a + b)$$

is positive, or if $\delta\theta$ is positive the expression within the brackets is positive. The value of $\alpha_1(\theta, \theta)$ at the value $\theta - \delta\theta$ is $\alpha_1(\theta, \theta) - \delta\theta(a + b)$, which is smaller than $\alpha_1(\theta, \theta)$.

Since we have assumed continuity of the functions involved, throughout a neighbourhood (over a square) around the point (θ, θ), $\alpha_1(\theta_1, \theta_2)$ lies between $\alpha_1(\theta, \theta) \pm \delta\theta(a + b)/2$. Consequently, throughout this square around (θ, θ), $\alpha_1(\theta_1, \theta_2)$ exceeds the value $\alpha_1(\theta - \delta\theta, \theta - \delta\theta)$ at the neighbouring point. It is clearly undesirable that more errors are committed when the populations are different than when they are equal in any given region including the line of equality (at least as a boundary) in which the possible values of (θ_1, θ_2) are restricted to lie. A necessary condition for this is that $(a + b)$ should vanish at all points on the line of equality, implying that $\alpha_1(\theta, \theta)$ and therefore $\alpha_2(\theta, \theta)$ should be constant independently of the common values.

We are not, at the moment, demanding that the functions $\alpha_1(\theta_1, \theta_2)$ and $\alpha_2(\theta_1, \theta_2)$ should be stationary or that they should be absolutely minimum on the line of equality, although these appear to be desirable properties leading to unbiased divisions of the space. It is, however, necessary that $\alpha_i(\theta, \theta)$ should be constant independently of θ. In our arguments, we have explicitely used one parameter although we said that θ stands for a vector of parameters. This is clearly admissible since we can consider variations in one parameter keeping the others fixed.

The restriction that $\alpha_i(\vartheta, \theta)$ is constant on the line of equality implies that with respect to the probability density of observations

$$P_1(x^1 \mid \theta)\, P_2(x^2 \mid \theta)\, p(x \mid \theta)$$

that is, when $\theta_1 = \theta_2 = \theta$, the regions R_1 and R_2 are similar to the sample space with respect to the free parameter θ. In such a case we shall say that there exists a ***similar division*** of the sample space with respect to θ.

Having determined similar divisions, we have to select the best one among

them. It is hard to imagine that there exist regions which minimise uniformly any linear compound of the errors $\pi_1\alpha_1(\theta_1, \theta_2) + \pi_2\alpha_2(\theta_1, \theta_2)$ except in some special cases. Some suitable criteria have to be used, as in Section 6, depending on the type of difficulties which the probability densities may present, to obtain reasonable solutions.

We have yet to consider the nature of the error functions on the line of equality where the maximum error for any group cannot be reduced below 50 per cent. It may be reasonable to impose the restriction

$$\alpha_1(\theta, \theta) = \alpha_2(\theta, \theta) = 0.50$$

In some problems the actual specification of the ratio of errors $\alpha_1(\theta, \theta) / \alpha_2(\theta, \theta)$ may be left open, and chosen to satisfy some optimum conditions. We could impose any other restriction specifying the ratio of errors at any value of the set (θ_1, θ_2).

A special case is the choice $\alpha_2(\theta, \theta) = 0.05$, which leads to a test of significance of the null hypothesis H_2, that the observed individual belongs to the second group, against the alternative that he belongs to the first group. This will be useful in further subdividing the regions R_1 and R_2 in such a way that some portions lead to more confident classifications, while other portions permit only provisional decisions. Further theory is developed in the examples considered in the next section.

The arguments of this section can be extended to the case of more than two alternative populations. The division of the space into k regions must be such that the error committed for any group remains constant whenever the populations are identical, whatever may be the common values of the parameters. As in the case of two populations, we may choose this constant to be $1/k$ for each of the alternative populations. Also, any ratio of these constants may be specified, or sometimes suitably determined. Problems of tests of significance may be considered in a similar way.

The general postulate laid down in this section can be used in the solution of a wide variety of problems in classification. For instance, the problem of the greater mean (Bahadur and Robbins, [1]) admits a neat solution once this condition is imposed.

6. Some optimum conditions and derivation of decision rules. It is known (Neyman, [10]) that similar regions can be constructed, when a set of sufficient statistics exist, by considering the relative probability density of the observations, given the set of sufficient statistics. Lehmann and Scheffé [8] have shown recently that when the parameters admit a minimal set of sufficient statistics such that no function depending on them has zero expectation (in which case the set is said to be complete) then all similar regions have Neyman's structure. That sufficient statistics possess this unicity property under some conditions has been formally demonstrated by the author [14]. In the illustrations considered in this paper, these results are used without proof.

If T stands for the complete set of sufficient statistics for θ, then we can write down the joint densities of the observations under H_1 and H_2 as

$$H_1: \quad \mathcal{P}_1(T \mid \eta, \delta)P_1(x^1, x^2, x \mid \eta, \delta, T) = \mathcal{P}_1(\eta, \delta)P_1(\eta, \delta),$$

$$H_2: \quad \mathcal{P}_2(T \mid \eta, \delta)P_2(x^1, x^2, x \mid \eta, \delta, T) = \mathcal{P}_2(\eta, \delta)P_2(\eta, \delta),$$

where P_1 and P_2 are relative probability densities of observations given T, and $\mathcal{P}_1$ and $\mathcal{P}_2$ are the densities of T, while η and δ are the vectors $(\theta_1 + \theta_2)$ and $(\theta_1 - \theta_2)$.

The regions R_{1T} and R_{2T} on the surface of T for which the linear compound of overall errors $a\alpha_1(\theta_1, \theta_2) + b\alpha_2(\theta_1, \theta_2)$ is a minimum subject to the condition

$$\alpha_1(\theta, \theta)/\alpha_2(\theta, \theta) = 1/\rho, \tag{6.1}$$

where ρ is fixed, are given by

$$R_{1T}: \quad a\mathcal{P}_1(\eta, \delta)P_1(\eta, \delta) + \lambda_1 P_1(\eta, 0) \geqq b\mathcal{P}_2(\eta, \delta)P_2(\eta, \delta) + \lambda_2 P_2(\eta, 0), \tag{6.2}$$

with the reverse relationship in R_{2T}. The constants λ_1 and λ_2 are determined to satisfy the condition (6.1). The proof of the result (6.2) and the subsequent ones follow from a lemma proved by the author in ([16], p. 340). The region (6.2) will generally depend upon the unknown quantities η and δ, and is therefore not useful. We therefore need to restrict the regions by imposing some condition on the error functions.

We first note that the errors $\alpha_1(\theta_1, \theta_2)$ and $\alpha_2(\theta_1, \theta_2)$ could be written in terms of η and δ as $\alpha_1(\eta, \delta)$ and $\alpha_2(\eta, \delta)$, using α_1 and α_2 as symbols for error functions. Let

$$\alpha_i'(\eta, \delta) = \frac{\partial}{\partial \delta}\alpha_i(\eta, \delta), \qquad i = 1, 2,$$

denote the derivatives with respect to the parameters δ in any given direction. The values $\alpha_1(\eta, 0)$ and $\alpha_2(\eta, 0)$ are the errors when the populations are identical and the slopes of the error functions in the given direction at $\delta = 0$ are

$$\alpha_1'(\eta, 0), \qquad \alpha_2'(\eta, 0). \tag{6.3}$$

To ensure optimum properties, at least in the neighbourhood of the line of equality of the two populations, we may minimise a linear compound of the slopes (6.3), or minimise them in a given ratio. Observing that minimising the slopes (6.3) is equivalent to minimising the slopes corresponding to the relative errors on the surfaces of T, we find the boundary separating the best regions R_{1T} and R_{2T} on the surfaces of T as

$$a\frac{\partial}{\partial \delta}P_1(\eta, 0) + \lambda_1 P_1(\eta, 0) = b\frac{\partial}{\partial \delta}P_2(\eta, 0) + \lambda_2 P_2(\eta, 0). \tag{6.4}$$

(i) For any a and b and the choice of λ_1 and λ_2 to satisfy the condition (6.1), the linear compound $a\alpha_1'(\eta, 0) + b\alpha_2'(\eta, 0)$ is minimised. The special values $a = b$ may be useful in practice.

(ii) For a suitable choice of a, b, λ_1, and λ_2, the slopes $\alpha_1'(\eta, 0)$ and $\alpha_2'(\eta, 0)$ can be minimised in a given ratio in addition to the condition (6.1) being satisfied. The special case of the equality of the slopes may be of some practical interest.

The solution (6.4) may depend on η when $P_1'(\eta, 0)$ and $P_2'(\eta, 0)$ contain η. In the illustrations considered in Section 7, the $P_i'(\eta, \delta)$ are functions of δ only, so that the solution (6.4) serves the purpose. Otherwise some method has to be devised, such as minimising the average slopes over a set of η or considering regions similar for η with respect to the functions $P_i'(\eta, 0)$.

For the problem of testing the hypothesis H_2 against the alternative H_1 we have to construct a region w on the T surfaces satisfying the four conditions (given $\gamma \leqq 0$)

$$(6.5)\quad \begin{cases} \text{(a)} & \int_w P_2(\delta = 0)\, dv = 0.05, \\ \text{(b)} & \int_w P_2'(\delta = 0)\, dv = \gamma, \\ \text{(c)} & \int_w P_2(\delta) \qquad dv \leqq 0.05, \\ \text{(d)} & \int_w P_1'(\delta = 0)\, dv = \text{a maximum}. \end{cases}$$

The region satisfying the conditions (a), (b) and (d) is given by

$$(6.6)\qquad w: \quad aP_1'(0) + \lambda_1 P_1(0) \geqq bP_2'(0) + \lambda_2 P_2(0)$$

on the T surfaces where a, b, λ_1 and λ_2 are suitably chosen. For this region the slope of the conditional power curve $\beta_1'(\delta)$ at $\delta = 0$ is a function of γ defined in condition (b). We now relax this condition and maximise $\beta_1'(0)$ subject to the condition $\gamma \leqq 0$. With this choice of γ we can set up the region w as in (6.6). If, for this region, condition (c) is satisfied, then we obtain a test of the hypothesis that H_2 is true against the alternative that H_1 is true. This test is most powerful in a given direction for small differences in the parameters of the two populations.

The situations in tests of significance and discriminatory problems are diagrammatically represented in Figure 1.

If the direction used in the above construction with the first derivatives is not justifiable, then we may try to impose further restrictions such as unbiasedness of the error functions on the line of equality

$$(6.7)\qquad \alpha_1'(\eta, 0) = 0, \qquad \alpha_2'(\eta, 0) = 0.$$

We will assume that this condition implies that the derivatives of these errors vanish when $\delta = 0$ for all T. The derivatives are calculated from the conditional

probability densities

$$\frac{\partial}{\partial \delta} \alpha_1(\eta, \delta, T), \qquad \frac{\partial}{\partial \delta} \alpha_2(\eta, \delta, T), \tag{6.8}$$

where $\int \alpha_i(\eta, \delta, T)\mathcal{P}_i(T \mid \eta, \delta)\, dT = \alpha_i(\eta, \delta)$. In all the illustrations considered in Section 7 this condition is automatically satisfied. Otherwise it may be necessary to impose the conditions (6.8) which may be only sufficient for (6.7).

We consider the second derivatives of the relative probability densities with respect to the elements of the vector of parameters $\delta = (\delta_1, \delta_2, \cdots)$. Defining for $k = 1$ or 2

$$P_k^{ij} = \frac{\partial^2}{\partial \delta_i\, \partial \delta_j} P_k(\delta = 0), \qquad P_k^i = \frac{\partial}{\partial \delta_i} P_k(\delta = 0),$$

let us construct the regions

$$\begin{aligned} (6.9) \quad R_1: \quad & \sum\sum a_{ij} P_1^{ij} + \lambda_{11} P_1^1 + \lambda_{12} P_1^2 + \cdots + \mu_1 P_1 \\ & \geqq \sum\sum b_{ij} P_2^{ij} + \lambda_{21} P_2^1 + \lambda_{22} P_2^2 + \cdots + \mu_2 P_2 , \end{aligned}$$

with the reverse relationship in R_2.

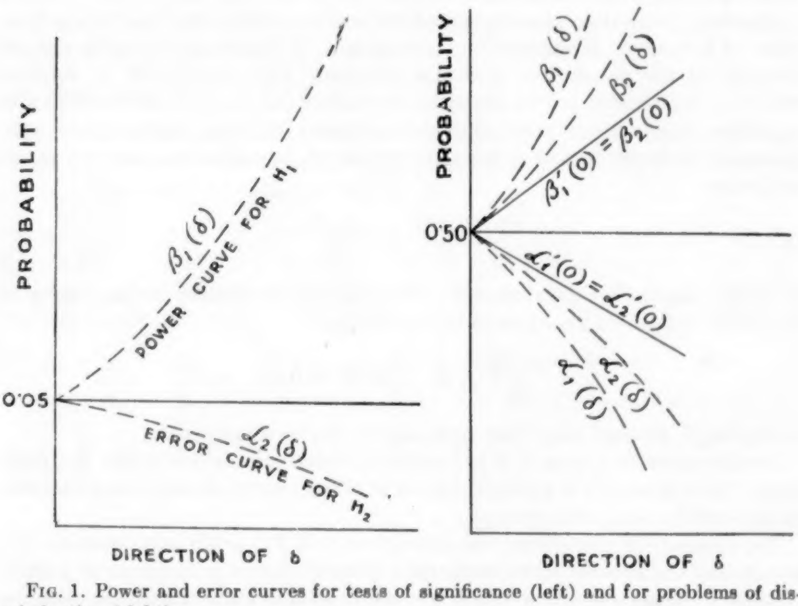

FIG. 1. Power and error curves for tests of significance (left) and for problems of discrimination (right).

(i) The regions R_1 and R_2 minimise

$$\sum\sum a_{ij} \frac{\partial^2}{\partial\delta_i \, \partial\delta_j} \alpha_1(\eta, \delta) + \sum\sum b_{ij} \frac{\partial^2}{\partial\delta_i \, \partial\delta_j} \alpha_2(\eta, \delta)$$

at $\delta = 0$ for given a_{ij} and b_{ij}, provided λ_{ij} and μ_i are chosen to satisfy the condition (6.8) and a given ratio of errors when $\delta = 0$.

(ii) For a suitable choice of a_{ij} and b_{ij}, the local powers of discrimination for the two groups can be made constant on the ellipses

$$\sum\sum \gamma^{ij}\delta_i\delta_j = \text{constant} \tag{6.10}$$

and their sum then maximised. Condition (6.10) implies that the second derivatives are in the ratio γ^{ij}.

(iii) By a suitable choice of a_{ij} and b_{ij} we could also construct a critical region w of a given size such that the first derivatives of $\alpha_1(\eta, \delta)$ and $\alpha_2(\eta, \delta)$ vanish at $\delta = 0$ and that $\int_w \sum\sum \gamma^{ij} P_1^{ij} \, dv$ is maximised subject to the condition $\int_w \sum\sum \beta^{ij} P_2^{ij} \, dv \leqq 0$, where γ^{ij} and β^{ij} are assigned as in (6.10). Such a region can be used in testing the hypothesis H_2 against the alternative H_1, provided the region is so adjusted that its size under H_2 is 5 per cent when $\delta = 0$ and less than or equal to 5 per cent when $\delta \neq 0$.

Another alternative is to restrict to those regions which give the errors as functions of a distance Δ between two populations. (Distance is a suitably defined function of the parameters of two populations. The construction of distance functions is discussed in two papers by the author [12], [14].) Even restricting to this class, it may not be possible to obtain regions for which a given linear compound of the errors $a\alpha_1(\Delta) + b\alpha_2(\Delta)$ is minimised. In such a case, we may try to minimise

$$a \frac{d\alpha_1(0)}{d\Delta} + b \frac{d\alpha_2(0)}{d\Delta} \tag{6.11}$$

to obtain regions for discrimination. For tests of significance we may have to maximise $-d\alpha_1(0)/d\Delta$ subject to the conditions

$$\frac{d\alpha_2(0)}{d\Delta} \leqq 0, \qquad \alpha_2(0) = 0.05.$$

In Section 7, we shall show that such regions can be constructed.

Another possible approach is to consider decision rules which satisfy the principle of invariance [7]. It appears that some of the results obtained here can also be deduced by using this principle.

The necessity of considering the derivatives in (6.11) arises only when no uniformly best regions exist in the class which gives the errors as functions of Δ only.

Besides the parameters θ which are considered to vary from population to population, there may be other unknown parameters ϕ which are the same for

all populations. Thus we may consider the class of normal distributions with the same unknown variance but different mean values. In such situations, we may demand that the division of the space be similar for the unknown parameters ϕ also when the populations are identical in the θ parameters.

This introduces fresh complications in the applications of the results (6.4), (6.6), (6.9) and (6.11) for the derivation of optimum regions. Fortunately, in some cases the problems can be reduced in such a way that the above results are directly applicable, as shown in Section 7.2.

One may argue that in laying down the decision rules, undue emphasis is laid on discriminating between populations which are close to one another. In the first place, this is done just to set up decision rules which do not involve the unknown parameters. In the absence of rules which are uniformly best, we can think only of rules which are best at some assigned values of the parameters, or at most for an assigned set of values.

The requirement that the decision rule should possess some optimum properties in the neighbourhood of equality of the populations is not unrealistic since in practice we often meet with alternatives which are closely related; the methods developed are best suited to such situations. It is, however, possible to reduce decision rules which have *optimum properties for a given difference in the parameters* of the two populations. These may be useful in some situations. Of course, whatever may be the rule offered, it is better to examine its performance for all possible differences in the parameters of the two populations and satisfy oneself whether it can be reasonably applied in a given situation.

7. Illustrations.

7.1. *Multivariate populations, dispersion matrix known.* Let us consider p characters and represent the relevant statistics computed from the three samples, and functions based on them, as follows:

Group	1	2	3
Sample size	n_1	n_2	n
Average of ith character	$\bar{x}_{i1}$	$\bar{x}_{i2}$	$\bar{x}_i$
Population average	μ_{i1}	μ_{i2}	μ_i

$$\delta_i = \mu_{i1} - \mu_{i2} \qquad Z_i = n\bar{x}_i + n_1\bar{x}_{i1} + n_2\bar{x}_{i2}$$

$$T_i = \bar{x}_{i1} - \bar{x}_{i2}, \qquad U_i = \bar{x}_i - (n_1\bar{x}_{i1} + n_2\bar{x}_{i2})/(n_1 + n_2)$$

$$f_1 = (n_1 + n_2)/n_1n_2, \qquad f_2 = (n + n_1 + n_2)/n(n_1 + n_2)$$

$$g_1 = n_2/(n_1 + n_2), \qquad g_2 = -n_1/(n_1 + n_2)$$

$$q_1 = 1/f_1 + g_1^2/f_1, \qquad q_2 = 1/f_2 + g_2^2/f_2$$

When $\mu_{i1} = \mu_{i2} = \mu_i$ for all i, then Z_i are sufficient statistics for μ_i and we need consider only the relative distribution of T_i, U_i given Z_i. It is easy to see that T_i, U_i, and Z_i are all independently distributed, so Z_i can be dropped from

further consideration if errors are restricted to functions of δ only. The joint probability density of T_i, U_i under the hypothesis H_1 is

$$P_1(T, U, \delta)$$
$$= \text{const}\exp\left[-\tfrac{1}{2}\sum\sum\alpha^{ij}\left\{\frac{(T_i-\delta_i)(T_j-\delta_j)}{f_1}+\frac{(U_i-g_1\delta_i)(U_j-\delta_j g_1)}{f_2}\right\}\right].$$

Under H_2, we replace g_1 by g_2 to obtain $P_2(T, U, \delta)$.

In this problem we consider regions R_1 and R_2 whose size under both hypotheses depends only on the single parameter $\Delta = \sum\alpha^{ij}\delta_i\delta_j$, since the formulae of Section 6, using the first and second derivatives, do not yield fruitful results. With this end in view let us consider the surface integral

$$\int_\Delta P_1(T, U, \delta)G^{-1}\,d\delta_1\cdots d\delta_p$$

where

$$\begin{aligned} &P_1(T, U, \delta)\\ (7.1.1)\qquad &= P_1(T, U, 0)\exp\{-\tfrac{1}{2}q_1\sum\sum\alpha^{ij}[(\delta_i - W_i)(\delta_j - W_j) - W_iW_j]\},\\ &G = \int_\Delta d\delta_1\cdots d\delta_p, \qquad W_i = (T_i/f_1 + g_1U_1/f_2)\div q_1.\end{aligned}$$

For the above integration, only the first expression in the exponential of (7.1.1) is important. This may be regarded as a p-variant normal distribution of $\delta_1, \cdots, \delta_p$. Then the integration results in a noncentral χ^2 probability density with noncentral parameter M_1, given by

$$(7.1.2)\qquad \left\{\frac{q_1^{p/2}\,|\,\alpha^{ij}\,|^{\frac{1}{2}}}{\pi^{p/2}}\right\}^{-1} e^{-M_1/2-\Delta'/2}(\Delta')^{(p/2)-1}\sum\left(\frac{M_1\Delta'}{4}\right)^r\frac{1}{r!\Gamma\left(\frac{p}{2}+r\right)}$$

$$M_1 = q_1\sum\sum\alpha^{ij}W_iW_j, \qquad \Delta' = q_1\Delta,$$

([16], pp. 51, 57). Observing that

$$\int_\Delta d\delta_1\cdots d\delta_p = \left\{\frac{|\,\alpha^{ij}\,|^{\frac{1}{2}}}{(2\pi)^{p/2}}\right\}^{-1}\frac{\Delta^{(p/2)-1}}{2^{p/2}\Gamma\left(\frac{p}{2}\right)}\,d\Delta$$

and changing over to Δ in (7.1.2), we find the total integral in (7.1.1) for the two cases to be

$$G_1(\Delta) = P_1(T, U, 0)e^{-q_1\Delta/2}\sum_0^\infty\frac{\Gamma(\frac{1}{2}p)}{r!\Gamma(\frac{1}{2}p+r)}\left(\frac{M_1q_1\Delta}{4}\right)^r,$$

$$G_2(\Delta) = P_2(T, U, 0)e^{-q_2\Delta/2}\sum_0^\infty\frac{\Gamma(\frac{1}{2}p)}{r!\Gamma(\frac{1}{2}p+r)}\left(\frac{M_2q_2\Delta}{4}\right)^r.$$

Restricting the minimisation of a linear compound of the errors to the divisions which yield errors as functions of Δ only, the boundary is obtained as

(7.1.3) $$aG_1(\Delta) = bG_2(\Delta).$$

The proof of (7.1.3) is trivial ([16], p. 285). We have to make sure that for the regions based on (7.1.3) the errors are functions of Δ only. This follows from the invariance of the expressions M_1 and M_2. The solution (7.1.3) in general involves Δ and can be used only when Δ is known. We can, however, seek for optimum properties in the neighbourhood of $\Delta = 0$, where

$$\frac{dG_1(\Delta)}{d\Delta} = P_1(T, U, 0)q_1\left(\frac{M_1}{2p} - \frac{1}{2}\right),$$

$$\frac{dG_2(\Delta)}{d\Delta} = P_2(T, U, 0)q_2\left(\frac{M_2}{2p} - \frac{1}{2}\right).$$

Consider the boundary

(7.1.4) $$a\,\frac{dG_1(0)}{d\Delta} + \lambda_1 P_1(T, U, 0) = b\,\frac{dG_2(0)}{d\Delta} + \lambda_2 P_2(T, U, 0),$$

or $aq_1M_1 - bq_2M_2 = c$. The choice $a = b$ leads to a minimum value of the sum of the derivatives of the errors. In this case the boundary is

(7.1.5) $$\sum\sum\alpha^{ij}\left\{\frac{g_1^2 - g_2^2}{f_2^2}\,U_iU_j + \frac{2(g_1 - g_2)}{f_1f_2}\,T_iU_j\right\} = \frac{p(g_1^2 - g_2^2)}{f_2}.$$

For the case $g_1 = -g_2$, equation (7.1.5) reduces to $\sum\sum\alpha^{ij}T_iU_j = 0$, so that the regions are

$$R_1: \quad \sum\sum\alpha^{ij}T_iU_j \geqq 0, \qquad R_2: \quad \sum\sum\alpha^{ij}T_iU_j \leqq 0,$$

with fifty per cent errors when $\Delta = 0$. In this case the regions are uniformly best for all Δ because $G_1(\Delta) \geqq G_2(\Delta)$ in R_1 and the reverse is true in R_2, irrespective of the value of Δ. The appropriate regions when $n_1 \neq n_2$ have the boundary as in (7.1.5). For these regions the errors may not be fifty per cent when $\Delta = 0$. If this condition is also insisted upon, the boundary is

(7.1.6) $$\sum\sum\alpha^{ij}\left\{\frac{g_1^2 - g_2^2}{f_2^2}\,U_iU_j + \frac{2(g_1 - g_2)}{f_1f_2}\,T_iU_j\right\} \geqq c,$$

where c is suitably determined. We can also choose a, b, λ_1, and λ_2 in (7.1.4) subject to the condition that the derivatives of errors are equal when $\Delta = 0$.

For tests of significance the critical region is of the form

(7.1.7) $$w: \quad aM_1 - bM_2 \geqq c$$

where a, b, and c are determined such that

(7.1.8) $$\frac{d}{d\Delta}\,\{P_1(aM_1 - bM_2 \geqq c)\}_{\Delta=0}$$

is a maximum subject to

$$P_2(aM_1 - bM_2 \geqq c \mid \Delta = 0) = 0.05, \qquad \frac{d}{d\Delta}\{P_2(aM_1 - bM_2 \geqq c)\}_{\Delta=0} \leqq 0.$$

In the above expressions, P_1 stands for the probability according to the first hypothesis and P_2 for the second. The ultimate solution depends on the evaluation of the expressions (7.1.8). The problem needs further investigation.

In the univariate problem, if $(n_1 - n_2)$ is not large compared to $(n_1 + n_2)$, the regions for classification are obtained as special cases of (7.1.5) as

$$R_1: \quad TU \geqq 0, \qquad R_2: \quad TU \leqq 0,$$

$$T = \bar{x}_1 - \bar{x}_2, \qquad U = \bar{x} - (n_1\bar{x}_1 + n_2\bar{x}_2)/(n_1 + n_2),$$

where $\bar{x}_1$ and $\bar{x}_2$ are the averages of the two samples and $\bar{x}$ that of the sample to be classified. The critical region for testing H_2 against H_1 is of the form $TU \geqq c$, where c is determined to ensure five per cent size when $\delta = 0$. The regions for classification depending on different combinations of T and U are diagrammatically represented in Figure 2.

7.2. *Multivariate populations, dispersion matrix unknown.* In addition to the statistics defined in Section 7.1 we need estimates of the dispersion elements when the populations are identical, that is, when $\delta_i = 0$ for all i. Let S_{ij} denote the pooled sum of products within the three samples with $(n_1 + n_2 + n - 3)$

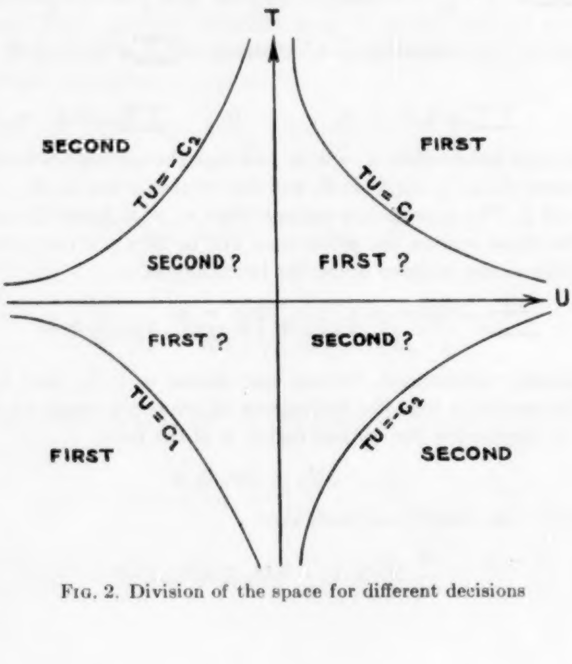

FIG. 2. Division of the space for different decisions

degrees of freedom. When δ_1 is zero, all the observations can be regarded as samples from the same population so that we have estimates of the dispersion elements based on $(n_1 + n_2 + n - 1)$ degrees of freedom. If B_{ij} denotes the corrected sum of products from the combined samples, then

$$B_{ij} = S_{ij} + T_iT_j/f_1 + U_iU_j/f_2 .$$

The statistics Z_i (defined in Sec. 7.1) and B_{ij} are sufficient for the common mean values and the elements α_{ij} of the dispersion matrix. Similar divisions of the space are obtained by considering exclusive regions on the surfaces of constant values of Z_i and B_{ij}, subject to some conditions. The probability density of T_i, U_i, and S_{ij} under the hypothesis H_1 is

$$\text{const } |S_{ij}|^{m/2} \exp \{-\tfrac{1}{2}\textstyle\sum\sum\alpha^{ij}[S_{ij} + (T_i - \delta_i)(T_j - \delta_j)/f_1 + (U_i - g_1\delta_i)(U_j - g_1\delta_j)/f_2]\}$$

where $m = (n_1 + n_2 + n - p - 4)$. Changing over to T_i, U_i, and B_{ij} permits their joint density to be written as the product of

$$P(B_{ij} \mid \delta = 0) = \text{const } |B_{ij}|^{m/2+1}, \exp \{-\tfrac{1}{2}\textstyle\sum\sum\alpha^{ij}B_{ij}\},$$

$$F(B, T, U) = |B_{ij} - T_iT_j/f_1 - U_iU_j/f_2|^{m/2} \div |B_{ij}|^{m/2+1},$$

$$Q_1(\delta) = \exp \{\textstyle\sum(T_i/f_1 + g_1U_i/f_2)\zeta_i - \tfrac{1}{2}\psi_1^2,$$

where $\zeta_1 = \alpha^{1i}\delta_1 + \cdots + \alpha^{pi}\delta_p$ and $\psi_1^2 = q_1\Delta = q_1\sum\sum\alpha_{ij}\zeta_i\zeta_j$. The probability density under the second hypothesis is obtained by replacing g_1 and q_1 by g_2 and q_2 in the above expressions. We shall consider divisions R_1, R_2 for which the errors are a function of the Mahalanobis distance $\Delta = \sum\sum\alpha_{ij}\zeta_i\zeta_j$ only. This means

$$(7.2.1) \qquad \int P(B_{ij} \mid \delta = 0)\, dB \int_{R_{1B}} e^{\psi_1^2/2}F(B, T, U)Q_1(\delta)\, dT\, dU = \beta_1(\Delta).$$

Following the arguments of Hsu [6] and Simaika [17] in a similar situation, we can show that condition (7.2.1) implies

$$\int_{R_{1B}} e^{\psi_1^2/2}F(B, T, U)Q_1(\delta)\, dT\, dU = G_1(K),$$

where $K = \sum\sum B_{ij}\zeta_i\zeta_j$. If we are minimising a linear compound of errors it is enough to maximise $ae^{-\psi_1^2/2}G_1(K) + be^{-\psi_2^2/2}G_2(K)$ on the surfaces of B_{ij}, since the expected value of this linear compound integrated over B_{ij} with density $P(B_{ij} \mid \delta = 0)$ gives the linear compound of correct classifications to be maximised. There is no hope of obtaining a solution without involving Δ, except perhaps when $n_1 = n_2$. We shall therefore minimise a linear compound of the derivatives with respect to Δ at the value zero, or maximise

$$(7.2.2.) \qquad a\frac{d}{d\Delta}\{e^{-\psi_1^2/2}\beta_1(\Delta)\} + b\frac{d}{d\Delta}\{e^{-\psi_2^2/2}\beta_2(\Delta)\}$$

$+ \Delta = 0$. Evaluation of the three terms at $\Delta = 0$ yields

$$e^{-\psi_1^2/2}\beta_1(\Delta) = \int P(B_{ij} \mid \delta = 0)\, dB \int_{R_{1B}} F(B, T, U) Q_1(\delta)\, dT\, dU,$$

$$= e^{-\psi_1^2} \int P(B_{ij} \mid \delta = 0) G_1(K)\, dB;$$

$$\alpha_{ij} \frac{d\beta_1(0)}{d\Delta} = \int B_{ij} P(B_{ij} \mid \delta = 0) \frac{dG_1(0)}{dK}\, dB,$$

$$= \frac{dG_1(0)}{dK} \int B_{ij} P(B_{ij} \mid \delta = 0)\, dB = (m + p + 3)\alpha_{ij} \frac{dG_1(0)}{dK};$$

$$\frac{d}{d\Delta} e^{-\psi_1^2/2}\beta_1(\Delta) = -\frac{q_1}{2}\beta_1(0) + \beta_1'(0).$$

Consequently the value of (7.2.2) at $\Delta = 0$ is

$$\left\{a \frac{dG_1(0)}{dK} + b \frac{dG_2(0)}{dK}\right\}(m + p + 3) - \left\{\frac{aq_1}{2} G_1(0) + \frac{bq_2}{2} G_2(0)\right\}.$$

Since the latter depends only on the errors when $\Delta = 0$, we need only maximise the former or the expression $a dG_1(0)/dK + b dG_2(0)/dK$ if possible, subject to given magnitudes of errors when $\Delta = 0$. Observing that

$$G_1(K) = \int_{R_{1B}} F(B, T, U) \exp\{\sum (T_i/f_1 + g_1 U_i/f_2)\zeta_i\}\, dT\, dU,$$

let us consider the surface integral over the surface $S = \sum\sum B_{ij}\zeta_i\zeta_j$

$$\int_S F(B, T, U) \exp\{\sum (T_i/f_1 + g_1 U_i/f_2)\zeta_i\} G^{-1}\, d\zeta_1 \cdots d\zeta_p, \tag{7.2.3}$$

$$G = \int_S d\zeta_1 \cdots d\zeta_p.$$

As in (7.1.1), the value of (7.2.3) is

$$\text{const } F(B, T, U) \sum \frac{\Gamma(\frac{1}{2}p)}{r!\Gamma(\frac{1}{2}p + r)} \left(\frac{M_1 K}{4}\right)^r, \tag{7.2.4}$$

where

$$M_1 = \sum\sum B^{ij}(T_i/f_1 + g_1 U_i/f_2)(T_j/f_1 + g_1 U_j/f_2).$$

The derivative of (7.2.4) with respect to K at $K = 0$ is const M_1, and the derivative corresponding to the second hypothesis is const M_2 where the two constants are the same. We can now define the boundary $aM_1 - bM_2 = c$ over the surfaces of B_{ij}. The constants a, b, and c may be suitably chosen. For discrimination we might choose $a = b$ and $c = 0$, in which case the sum of the derivatives of the errors is minimised. We can choose a, b, and c differently as in other cases considered in Section 7.1.

For tests of significance we need to determine a, b, and c such that over the surfaces of B_{ij}

$$\int_w F(B, T, U)Q_2(0)\, dT\, dU = 0.05 \qquad \frac{d}{dK}\int_w F(B, T, U)Q_1(\delta)\, dT\, dU$$

is a maximum at $K = 0$ subject to

$$\frac{d}{dK}\int_w F(B, T, U)Q_2(\delta)\, dT\, dv \leqq 0, \qquad k = 0.$$

Here w is the region on the surfaces of B_{ij} where $aM_1 - bM_2 \geqq c$. It is easy to see that the distribution of the statistic $aM_1 - bM_2$ under any hypothesis is dependent on Δ only, thus ensuring the validity of the arguments used in the derivation of the regions.

The distribution problems connected with the test criteria developed here have yet to be tackled. Some results obtanied by Wald [22], Harter [5] and Sitgreaves [18] in the reduction of distribution of the discriminant function and difference of two quadratic forms will be extremely useful in the study of these problems.

REFERENCES

[1] R. R. Bahadur and H. Robbins, "The problem of the greater mean," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 469–487.

[2] R. A. Fisher, "The fiducial argument in statistical inference," *Ann. Eugenics*, Vol. 6 (1935), pp. 391–398.

[3] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, Vol. 7 (1936), pp. 179–188.

[4] E. Fix and J. L. Hodges, "Discriminatory analysis: nonparametric discrimination: consistency problems", School of Aviation Medicine. Project number 21-49-004, (1951).

[5] H. L. Harter, "On the distribution of Wald's classification statistic," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 58–67.

[6] P. L. Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), pp. 62–69.

[7] E. L. Lehmann, "Theory of testing hypotheses," (Lecture notes, 1948).

[8] E. L. Lehmann and H. Scheffe, "Completeness, similar regions and unbiased estimation," *Sankhya*, Vol. 10 (1950), pp. 305–340.

[9] R. v. Mises, "On the classification of observation data into distinct groups," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 68–73.

[10] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philos. Trans. Roy. Soc. London, Ser. A*, Vol. 236 (1937), pp. 333–380.

[11] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1-161.

[12] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81–91.

[13] C. R. Rao, "The utilization of multiple measurements in problems of biological classification," *J. Roy. Stat. Soc. Suppl.*, Vol. 10 (1948), pp. 159–193.

[14] C. R. Rao, "Sufficient statistics and minimum variance estimates," *Proc. Cambridge Phil. Soc.*, Vol. 45 (1949), pp. 213–218.

[15] C. R. RAO, "Statistical inference applied to classificatory problems," *Sankhya*, Vol. 10 (1950), pp. 229–256, Vol. 11 (1951), pp. 107–116, 257–272, Vol. 12 (1952–53), pp. 229–246.

[16] C. R. RAO, "*Advanced Statistical Methods in Biometric Research*," John Wiley and Sons, 1952, New York.

[17] J. B. SIMAIKA, "On the optimum property of two important statistical tests," *Biometrika*, Vol. 32 (1941), pp. 70–80.

[18] R. SITGREAVES, "On the distribution of two random matrices used in classification procedures," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 263–270.

[19] A. WALD, "Contributions to the theory of statistical estimation and testing hypotheses," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 299–326.

[20] A. WALD, "Asymptotically most powerful tests of statistical hypotheses," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 1–19.

[21] A. WALD, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.

[22] A. WALD, "On a statistical problem arising in the classification of an individual into one of two groups," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 145–162.

[23] A. WALD, "Statistical decision functions," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 165–205.

[24] B. L. WELCH, "Note on discriminant functions," *Biometrika*, Vol. 31 (1939), pp. 218–220.

A CONFIDENCE INTERVAL FOR VARIANCE COMPONENTS

By J. R. Green

University of Liverpool

1. Introduction.

Summary. In this paper an approximate confidence interval is found for the expected value of the difference between two quantities which are independently distributed proportionally to χ^2 variates. Three methods are used. The first is based on the work of Welch [13], [14] and Aspin [1], [2] on the generalized "Student's" problem, and involves neglecting successively higher powers of the reciprocal of one of the degrees of freedom. This method is used to check the other two solutions, both of which involve neglecting successive increasing and decreasing powers, respectively, of a nuisance parameter. Finally a solution is formed using those resulting from the second and third methods, and is more accurate than those solutions. The order of accuracy, and the use of the final solution, are discussed.

The paper does not present a method of computing confidence intervals in a form suitable for immediate *practical* application. Series developments of a certain hypothetical function are given; more remains to be said about the relation between the series and the function, and the problem of computing tables. A computational exploration of the solution is at present in hand.

Applications. In what has sometimes been termed a Model II multiple classification, each observation is the sum of a constant and of contributions due to the different factors which feature in the classification, the interaction effects, and an error term. These contributions are taken to be normally and independently distributed with zero means, and variances independent of the particular levels of the appropriate factors. These variances are called variance components, since each gives that portion of the total variance of each observation appropriate to a particular source. In a balanced layout, each of these variance components, except that due to the error term, can be written as a known constant multiplied by the difference of the expected values of two mean squares, which are independently distributed proportionally to χ^2 variates. Thus the results of this paper may be applied to these variance components.

In the other main model of multiple classifications, the so-called Model I, the factors make constant contributions to the observations at the different levels. Here all the mean squares except the residual are proportional to noncentral χ^2 variates, so that the results of this paper cannot be applied. However, in a "Mixed Model", where some factors are as in Model I and some as in Model II, some mean squares will be suitable.

The general balanced Model II classification will be exemplified by considering the two-way layout. Let y_{ijk} be the kth observation in the ith row and the jth

Received 11/16/53, revised 5/28/54.

column, and take $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, where μ is a constant, and α_i, β_j, γ_{ij}, and ϵ_{ijk} are independent normal deviates with variances σ_α^2, σ_β^2, σ_γ^2 and σ_ϵ^2, respectively. The appropriate table is thus

Source	D. F.	Mean square	$\mathcal{E}$ (mean square)
Between rows	$a - 1$	M_α	$\sigma_\epsilon^2 + n\sigma_\gamma^2 + nb\sigma_\alpha^2$
Between cols.	$b - 1$	M_β	$\sigma_\epsilon^2 + n\sigma_\gamma^2 + na\sigma_\beta^2$
Interaction	$(a - 1)(b - 1)$	M_γ	$\sigma_\epsilon^2 + n\sigma_\gamma^2$
Error	$ab(n - 1)$	M_ϵ	σ_ϵ^2

Consider the variance component σ_α^2, for example. Now

$$\sigma_\alpha^2 = (nb)^{-1}\mathcal{E}(M_\alpha - M_\gamma),$$

and M_α and M_γ are independently distributed as

$$(a - 1)^{-1}(\sigma_\epsilon^2 + n\sigma_\gamma^2 + nb\sigma_\alpha^2)\chi^2, \qquad (b - 1)^{-1}(\sigma_\epsilon^2 + n\sigma_\gamma^2)\chi^2,$$

so that the results of this paper may be applied to obtain confidence limits for σ_α^2.

It is well-known that a confidence interval or confidence limit can be used to provide a test of a hypothesis which postulates a particular value for the parameter concerned; for if the hypothesis be accepted when the hypothetical value of the parameter lies inside the interval or on the appropriate side of a single confidence limit, then the probability of rejecting a true hypothesis is fixed at some chosen level. This is true for the limits found for K, the variance component for which an interval estimate is obtained. An examination of the power of this test may require the tabulation of the function derived in Section 7. However it seems reasonable to expect that, for a sufficiently large difference between the true and hypothetical values of K, the power of the test will be a monotonically increasing function of this difference, since the interval continues to cover the true value in the fixed proportion of cases.

Crump [5] states three main fields of application of work on variance components.

(i) *The interpretation of significance tests.* Here variance component estimates are used to locate the sources of undesirable variation, so that this variation can be partially or completely eliminated. Tippett [12] discusses significance tests in the analysis of variance in terms of these components, giving a numerical example of the quality control of spectacle glass. Daniels [8] gives an example from the woollen industry.

(ii) *The selection of efficient sampling designs.* This is the most important use of variance components. Usually interest is focussed on a particular function of the observations, such as the grand mean. The reciprocal of the variance of this statistic is then regarded as a measure of the efficiency of the sampling design. Both the cost and the efficiency are functions of the sample sizes and variance components. The usual procedure for choosing a good design is to estimate the variance components from a preliminary experiment, and then, using these es-

timates instead of the true values, to calculate the sample sizes which either minimise the cost for fixed efficiency or maximise the efficiency for fixed cost. Alternatively, Yates [15] suggests the general principle that an experiment should be so designed that the sum of the cost and the expected losses due to errors in the results should be minimised. Examples are given by Marcuse [10], and Nordskog and Crump [6].

(iii) *Various problems in genetics.* An example is given by Robinson and Comstock [4].

In all of these fields, point estimates of variance components are now used. They seem to be more appropriate than interval estimates for many of the examples met in practice. However, a confidence interval is useful for assessing the accuracy of an estimate. If the confidence interval is wide, then little trust can be placed in a point estimate; if it is narrow, then the estimate can reasonably be regarded as trustworthy. Estimates do exist for the variances of the variance component estimates, but these, being estimates, are less reliable than confidence intervals for assessing the accuracy of the variance component estimates. Also, they are less informative, since the usual type of variance component estimate has a complicated distribution, involving a nuisance parameter (see K. Pearson [11]).

When variance components are used qualitatively to assess the amount of variation present, a confidence interval may be a more reliable guide to the judgement.

Previous work. A full discussion of previous work would require too much space and anything brief is scarcely illuminating. Attention is directed to papers by Fisher [9], Bross [3], and a comprehensive survey by Crump [7]. In this paper we do not follow Fisher's method of computing fiducial limits.

2. The problem. The previous problems may be subsumed under the following canonical form. Two statistics M_1 and M_2 are given, which are independently distributed as $\sigma_1^2\chi^2/r_1$ and $\sigma_2^2\chi^2/r_2$, with r_1 and r_2 degrees of freedom, respectively. Confidence limits are required for $\sigma_1^2 - \sigma_2^2$, both σ_1^2 and σ_2^2 being unknown. For the present, it will be assumed that $\sigma_1^2 > \sigma_2^2$, but this restriction will be withdrawn later, as discussed in Section 9.

We define

$$K = \sigma_1^2 - \sigma_2^2, \qquad \rho = \frac{\sigma_2^2}{K} = \frac{1}{(\sigma_1/\sigma_2)^2 - 1}, \qquad y = \frac{M_1}{K}, \qquad x = \frac{M_2}{K}.$$

Thus a function f is sought such that

$$\Pr\,[y \leqq f(x)] = \alpha, \qquad 0 < \alpha < 1, \tag{2.1}$$

where α is given and f must be independent of the nuisance parameter. Later we shall require to find K such that $M_1/K = f(M_2/K)$. The problem was put into this form originally so that the method of approach due to Welch (later referred to as Method I) might be exploited.

Now r_2x/ρ and $r_1y/(1 + \rho)$ are independently distributed as χ^2 on r_1 and r_2

degrees of freedom, respectively. With $y_1 = r_1 y/(1 + \rho)$, requirement (2.1) becomes

$$(2.2)\quad \int_0^\infty \frac{e^{-r_2 x/2\rho}}{\Gamma(\frac{1}{2}r_2)} \left(\frac{r_2 x}{2\rho}\right)^{r_2/2-1} \left\{\int_0^{r_1 f(x)/(1+\rho)} \left(\frac{y_1}{2}\right)^{r_1/2-1} \frac{e^{-y_1/2}}{2\Gamma(\frac{1}{2}r_1)}\, dy_1\right\} \frac{r_2\, dx}{2\rho} = \alpha.$$

Since x and y are nonnegative, the discussion is confined entirely to the first quadrant of the plane of x and y. Thus it is essential that $f(x) \geqq 0$ (see Section 9). Put

$$I_{r_1}(z) = \int_0^z \{(\tfrac{1}{2}y_1)^{r_1/2-1} e^{-y_1/2}/2\Gamma(\tfrac{1}{2}r_1)\}\, dy_1, \qquad g(x) = I_{r_1}\{r_1 f(x) \,/\, (1 + \rho)\}.$$

Further, let ξ be such that $I_{r_1}(\xi) = \alpha$. Thus an f is required such that

$$(2.3)\quad \int_0^\infty \frac{e^{-r_2 x/2\rho}}{\Gamma(\frac{1}{2}r_2)} \left(\frac{r_2 x}{2\rho}\right)^{r_2/2-1} \frac{r_2 g(x)}{2\rho}\, dx = \alpha = I_{r_1}(\xi),$$

independently of ρ. We do not know whether there exists a function $f(x)$ which satisfies the above conditions nor, if it exists, whether it is regular. In this paper we derive a function $f_{IV}(x)$ which is such that (2.2) is approximately satisfied when $f_{IV}(x)$ is inserted in place of $f(x)$. How good this approximation is can be determined only by computational means.

3. Method I. In equation (2.3) we expand $g(x)$ in a Taylor series about $x = \rho$. That is, we confine the investigation to the finding of a solution for which this is permissible, if one exists. Now $g(x) = e^{(x-\rho)\partial} g(w)$, where $\partial^r = [\partial^r/\partial w^r]_{w=\rho}$. Thus (2.3) becomes

$$\int_0^\infty \frac{e^{-r_2 x/2\rho}}{\Gamma(\frac{1}{2}r_2)} \left(\frac{r_2 x}{2\rho}\right)^{r_2/2-1} e^{(x-\rho)\partial} \frac{r_2\, dx}{2\rho}\, g(w) = \alpha.$$

With $\Theta = (1 - 2\rho\partial/r_2)^{-r_2/2} e^{-\rho\partial}$, this becomes $\Theta I_{r_1}\{r_1 f(w) \,/\, (1 + \rho)\} = \alpha$. Expansion of $I_{r_1}\{r_1 f(w) \,/\, (1 + \rho)\}$ about $\{r_1 f(w) \,/\, (1 + \rho)\} = \xi$ yields

$$I_{r_1}\left\{\frac{r_1 f(w)}{1 + \rho}\right\} = \exp\left[\left\{\frac{r_1 f(w)}{1 + \rho} - \xi\right\} D\right] I_{r_1}(z), \qquad D^r = \left[\frac{d^r}{dz^r}\right]_{z=\xi}.$$

Hence the equation to be solved becomes

$$(3.1)\quad \Theta \exp\left(\{[r_1 f(w) \,/\, (1 + \rho)] - \xi\} D\right) I_{r_1}(z) = \alpha.$$

Equation (2.2) is very similar to the one solved approximately by Welch [13], [14] and Aspin [1], [2] in their work on the problem of comparing two means; the method used here is the same as theirs. The different functional form of the inner integrand's upper limit, and the different type of the inner integral, prevent deriving our solution from Welch's, although a comparison means of checking for $r_1 = 1$ can be used. It is evident that Θ is essentially the same in both cases.

Continuing, we put $f = f_0 + f_1 + f_2 + \cdots$, where f_s is of order $-s$ in r_2, and the expansion may be finite or infinite. The quantity f_0 is, as in Welch's

work, the large sample approximation, here $\xi(1 + x)/r_1$. Expanding,

$$
\begin{aligned}
\Theta &= \exp\{-\rho\partial - \tfrac{1}{2}r_2 \log(1 - 2\rho\partial/r_2)\} \\
&= \exp\{\rho^2\partial^2/r_2 + 4\rho^3\partial^3/3r_2^2 + \cdots\} \qquad (3.2)\\
&= 1 + \rho^2\partial^2/r_2 + \{4\rho^3\partial^3/3r_2^2 + \rho^4\partial^4/2r_2^2\} + \cdots.
\end{aligned}
$$

Neglecting terms of order r_2^{-3} in (3.1), we have

$$\Theta \exp\left\{\xi D\left[\frac{1+w}{1+\rho} - 1\right]\right\}\left\{1 + \frac{r_1 f_1(w)D}{1+\rho} + \left[\frac{r_1 f_2(w)D}{1+\rho} + \frac{r_1^2 f_1^2(w)D^2}{2(1+\rho)^2}\right] + \cdots\right\} \cdot I_{r_1}(z) = I_{r_1}(\xi).$$

Substituting for Θ, and grouping separately terms of order r_2^{-1} and r_2^{-2}, we obtain

$$
\begin{aligned}
&\left[\frac{r_1 f_2(\rho)}{1+\rho} D + \frac{r_1 f_1^2(\rho)D^2}{2(1+\rho)^2} + \frac{\rho^2\partial^2}{r_2} \exp\left\{\xi D\left(\frac{1+w}{1+\rho} - 1\right)\right\} \frac{r_1 f_1(w)D}{1+\rho}\right.\\
&\qquad \left. + \left\{\frac{4\rho^3\partial^3}{3r_2^2} + \frac{\rho^4\partial^4}{2r_2^2}\right\} \exp\left\{\xi D\left(\frac{1+w}{1+\rho} - 1\right)\right\}\right] I_{r_1}(z)\\
&\qquad + \left[\frac{r_1 f_1(\rho)}{1+\rho} D + \frac{\rho^2\partial^2}{r_2} \exp\left\{\xi D\left(\frac{1+w}{1+\rho} - 1\right)\right\}\right] I_{r_1}(z) = 0.
\end{aligned}
$$

Equating to zero the first order term yields

$$[r_1 f_1(\rho) / (1+\rho)]I'_{r_1}(\xi) + [\rho^2\xi^2 / r_2(1+\rho)^2]I''_{r_1}(\xi) = 0.$$

Therefore $f_1(\rho) = -\rho^2\xi^2 I''_{r_1}(\xi) / r_1 r_2(1+\rho)I'_{r_1}(\xi)$. We put $R_s = I^{(s)}_{r_1}(\xi) / I'_{r_1}(\xi)$, so that

$$f_1(x) = -x^2\xi^2 R_2 / r_1 r_2(1+x).$$

Equating to zero the second order term yields

$$f_2(x) = \frac{-x^2\xi}{6r_1 r_2^2(1+x)^3}\{x^2[3R_2(\xi^3R_2^2 - 2\xi^3R_3 - 4\xi^2R_2) + 8\xi^2R_3 + 3\xi^3R_4] + 8x[\xi^2R_3 - 3\xi^2R_2^2] - 12\xi R_2\}.$$

It is required to express R_s in terms of ξ and r_1. Now

$$
\begin{aligned}
I_{r_1}(\xi) &= \int_0^\xi (\tfrac{1}{2}y)^{r_1/2-1}[e^{-y/2}/2\Gamma(\tfrac{1}{2}r_1)]\,dy,\\
I'_{r_1}(\xi) &= (\tfrac{1}{2}\xi)^{r_1/2-1}[e^{-\xi/2}/2\Gamma(\tfrac{1}{2}r_1)].
\end{aligned}
$$

Using Leibniz's formula,

$$I^{(s)}_{r_1}(\xi) = 2^{-s+1}\frac{(\tfrac{1}{2}r_1 - 1)!}{(\tfrac{1}{2}r_1 - s)!}\left(\frac{\xi}{2}\right)^{r_1/2-s}\frac{e^{-\xi/2}}{2\Gamma(\tfrac{1}{2}r_1)}$$

$$- \binom{s-1}{1} \frac{(\frac{1}{2}r_1 - 1)!}{(\frac{1}{2}r_1 - s + 1)!} 2^{-s+1} \left(\frac{\xi}{2}\right)^{r_1/2-s+1} \frac{e^{-\xi/2}}{2\Gamma(\frac{1}{2}r_1)}$$

$$+ \cdots + (-2)^{-s+1} \left(\frac{\xi}{2}\right)^{r_1/2-1} \frac{e^{-\xi/2}}{2\Gamma(\frac{1}{2}r_1)}$$

$$= 2^{-s} \left(\frac{\xi}{2}\right)^{r_1/2-s} \frac{e^{-\xi/2}}{\Gamma(\frac{1}{2}r_1)} \sum_{i=0}^{s-1} \left(\frac{-\xi}{2}\right)^i \frac{(\frac{1}{2}r_1 - 1)!}{(\frac{1}{2}r_1 - s + i)!} \binom{s-1}{i}.$$

Therefore

$$R_s = (2\xi)^{-s+1} \sum_{i=0}^{s-1} 2^{1+s-i} (-\xi)^i \frac{(\frac{1}{2}r_1 - 1)!}{(\frac{1}{2}r_1 - s + i)!} \binom{s-1}{i}$$

$$= (2\xi)^{-s+1} \Big\{ [r_1 - 2][r_1 - 4] \cdots [r_1 - 2(s-1)] - \binom{s-1}{1}$$

$$\cdot \xi([r_1 - 2] \cdots [r_1 - 2(s-2)]) + \cdots + (-\xi)^{s-1} \Big\}.$$

Substituting for R_s in the f's, we obtain

$$f_0(x) = (1 + x)\frac{\xi}{r_1}, \qquad f_1(x) = \frac{x^2}{1+x} \frac{\xi(\xi - r_1 + 2)}{2r_1 r_2}$$

$$f_2(x) = \frac{x^2\xi}{24r_1 r_2^2(1+x)^3} \{x^2[4\xi^2 - 11\xi(r_1 - 2) + (r_1 - 2)(7r_1 - 10)]$$

$$+ 16x[\xi^2 - 2\xi(r_1 - 2) + (r_1 - 1)(r_1 - 2)] + 24(r_1 - 2 - \xi)\}.$$

For further terms operate on

$$I_{r_1}\left\{\frac{f_0(w) + \cdots + f_r(w)}{1 + \rho} r_1\right\}$$

by $-\Theta$ and arrange the result as a power series in $1/r_2$, say $\sum a_{r+1}(\rho)/r_2^{r+1}$. Then

$$r_1 f_{r+1}(\rho) I_{r_1}(\xi)/(1 + \rho) = a_{r+1}(\rho)/r_2^{r+1}$$

whence $f_{r+1}(\rho)$. Now this expansion is in descending powers of r_2, but though this may be large compared with r_1 it may not be large compared with certain powers of r_1 which may occur in the numerators of the f's, or compared with ξ. This matter is considered in Section 8.

Only the terms shown above have been worked out by this method, as the calculations become laborious and this solution is used only as a check on the solutions obtained by other methods.

Another point regarding this solution is that in replacing ρ by x it has been assumed that a solution $f(x)$ exists which is independent of ρ, whereas the function $f(w)$ which is operated upon by ∂ may be actually of the form $f(w, \rho)$. However, if such a solution exists, then this method will give it. Moreover, $f = f_0 + f_1 + f_2$ does satisfy (2.2) to the order r_2^{-2}, whether or not an exact solution exists.

A check has been made in the case $r_1 = 1$, when it is possible to deduce the appropriate series from Welch's solution of the two means problem.

4. Graphical representation. At this stage, a picture helps one to visualize Methods II and III, described in the subsequent sections. For simplicity, put

$$v = \tfrac{1}{2}r_2x, \qquad u = \tfrac{1}{2}r_1y, \qquad a = \tfrac{1}{2}r_1 - 1, \qquad b = \tfrac{1}{2}r_2 - 1.$$

The joint probability density function of u and v is

$$\frac{1}{a!b!}\left(\frac{u}{1+\rho}\right)^a\left(\frac{v}{\rho}\right)^b\frac{1}{\rho(1+\rho)}\exp\left\{-\left[\frac{u}{1+\rho}+\frac{v}{\rho}\right]\right\},$$

where $a! = \Gamma(a+1)$ whether a is an integer or not. It is required to find $g(u, v)$ such that the integral of this density function over the region $g(u, v) \leqq 0$ shall equal α, that is, such that

$$(1/a!b!)\iint_{\mathfrak{D}} u^a v^b e^{-(u+v)}\,du\,dv = \alpha, \qquad \mathfrak{D} = \{u, v : g[u(1+\rho), v\rho] \leqq 0\}.$$

This equation shows that g is required such that when the curve $g(u, v) = 0$ is scaled down by dividing the u-value of every point by $(1 + \rho)$, and the v-value by ρ, then the integral of $(1/a!b!)\,u^a v^b e^{-(u+v)}$ over the region on one side of the resulting curve is α, independently of ρ. Also, when $g(u, v) = 0$, only one value of v corresponds to one of u.

When $\rho \to 0$, the slope becomes very small and the curve flattens out to the form $v =$ constant. If the integral under the curve is α, then the constant value of v will be $\tfrac{1}{2}\xi$, where ξ is such that $I_{r_1}(\xi) = \alpha$.

When $\rho \to \infty$, the curve cannot lie completely in the range $u \leqq U$ for any finite U. If it did, the scaled curve would lie completely in the range $u \leqq U/(1 + \rho)$, which becomes arbitrarily small as $\rho \to \infty$. Thus the integral on one side of the curve can be made arbitrarily small, or close to 1, as the case may be. Similarly the curve cannot lie wholly in the region $v \leqq V$, for any finite V. Thus the curve extends to infinity in both variables.

Further, looking for a solution whose slope tends to a definite limit (not necessarily finite) for large u and v (if such a solution exists), then for ρ large the shape of the scaled curve will be roughly that of a straight line through the origin, with slope equal to the slope at infinity. Now the integral on one side of this line, say below it, must be α, that is, Pr $(y/x \leqq m/r_1) = \alpha$. Since y/x is distributed as F_{r_1, r_2}, (i.e. the F variate with degrees of freedom r_1, r_2),

$$m = r_1F_{r_1, r_2}(\alpha), \qquad \text{where Pr}\,[(F_{r_1, r_2} \leqq F_{r_1, r_2}(\alpha)] = \alpha.$$

From this crude picture of the approximations for ρ very large and very small, the first stages of the approximations derived in Methods II and III can be obtained. A greater understanding of these methods is also provided.

The same conclusions are reached by the following intuitive reasoning:

When $\rho \to 0$, then $\rho = \sigma_2^2/\mathrm{K}$ and $\mathrm{K} = \sigma_1^2 - \sigma_2^2$. Let $\sigma_2^2 \to 0$, then $\mathrm{K} = \sigma_1^2$, so

that M_1 tends to become an unbiased estimate of K, distributed as $r_1^{-1}K\chi^2$ on r_1 degrees of freedom. Hence r_1y is distributed as χ^2 on r_1 degrees of freedom. An approximate $f(x)$ such that Pr $[r_1y \leqq r_1f(x)] = \alpha$ is $f(x) = \xi/r_1$, which is thus the limiting solution as $\rho \to 0$.

When $\rho \to \infty$, let $K \to 0$ and $\sigma_2^2 \to \sigma_1^2$, so that M_1/M_2 tends to become an F_{r_1,r_2} variate. Thus in the limit y/x is an F_{r_1,r_2} variate, and the appropriate $f(x)$ such that Pr. $[y \leqq f(x)] = \alpha$ is mx/r_1 which is thus the limiting solution as $\rho \to \infty$.

5. Method II. As $\rho \to 0$, M_1 tends to become an unbiased estimate of K, so that for ρ small, $f(x) = \xi/r_1$ would approximately satisfy (2.2), as just pointed out in the preceding section. This solution neglects terms of order ρ^2, so that the accuracy could be improved by neglecting higher orders of ρ instead.

We take $b = \frac{1}{2}r_2 - 1$ as before, and change (2.2), by the transformations of Section 4, to the form

$$\int_0^\infty \frac{e^{-u}u^b}{b!} I_{r_1}\left\{\frac{r_1f(2\rho u/r_2)}{1+\rho}\right\} du = \alpha. \tag{5.1}$$

It can be seen easily that it is appropriate to seek an $f(x)$ of the form

$$r_1f(x) = b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots.$$

We now expand the function I_{r_1} in (5.1) about the point where its argument equals ξ. When $\rho = 0$, we find that $b_0 = \xi$. The expansion gives

$$I_{r_1}\left\{\frac{b_0 + b_1(2\rho u/r_2) + b_2(2\rho u/r_2)^2 + \cdots}{1+\rho}\right\} = \alpha + \frac{p}{1!}I'_{r_1}(\xi) + \frac{p^2}{2!}I''_{r_1}(\xi) + \cdots,$$

where

$$\begin{aligned} p &= \{[b_0 + b_1(2\rho u/r_2) + b_2(2\rho u/r_2)^2 + \cdots]/[1+\rho]\} - \xi \\ &= [\rho(2b_1u/r_2 - \xi) \quad + b_2(2\rho u/r_2)^2 + \cdots]/[1+\rho]. \end{aligned}$$

Substitution in (5.1) gives

$$\int_0^\infty [e^{-u}u^b/b!][pI'_{r_1}(\xi) + (p^2/2!)I''_{r_1}(\xi) + \cdots]\, du = 0.$$

Dividing through by $I'_{r_1}(\xi)/(1+\rho)$ gives

$$\begin{aligned} \int_0^\infty \frac{e^{-u}u^b}{b!} [\{\rho(2b_1u/r_2 - \xi) + b_2(2\rho u/r_2)^2 + \cdots\} + \{R_2/2(1+\rho)\} \\ \{\rho(2b_1u/r_2 - \xi) + b_2(2\rho u/r_2)^2 + \cdots\}^2 + \{R_3/3!(1+\rho)^2\} \\ \{\rho(2b_1u/r_2 - \xi) + \cdots\}^3 + \cdots]\, du = 0. \end{aligned}$$

Equating to zero the coefficient of ρ gives $2b_1(b+1) = r_2\xi$, or $b_1 = \xi$. Similarly equating to zero the coefficient of ρ^2 gives

$$(2/r_2)^2b_2(b+2)(b+1) = -\tfrac{1}{2}R_2\{(2b_1/r_2)^2(b+2)(b+1) + \xi^2 - (4b_1/r_2)(b+1)\xi\}$$

Therefore $b_2(b + 2)/(b + 1) = -\frac{1}{2}R_2\xi^2\{(b + 2)/(b + 1) + 1 - 2\}$, and consequently $b_2 = -\frac{1}{2}R_2\xi^2/(b + 2)$.

Similar consideration of the coefficients of ρ^3 and ρ^4 gives, in turn,

$$b_3 = \frac{(b + 1)}{(b + 2)(b + 3)}\left\{\frac{R_2^2\,\xi^3}{b + 1} + \frac{R_2\,\xi^2}{2} - \frac{R_3\,\xi^3}{3(b + 1)}\right\},$$

$$b_4 = \frac{(b + 1)^2}{(b + 2)(b + 3)(b + 4)}\left\{\frac{(b + 11)R_2\,R_3\,\xi^4}{4(b + 1)^2} - \frac{5R_2^2\,\xi^3}{2(b + 1)} + \frac{2R_3\,\xi^3}{3(b + 1)}\right.$$
$$\left. - \frac{R_2\,\xi^2}{2} - \frac{(b^2 + 31b + 60)R_2^3\,\xi^4}{8(b + 1)^2(b + 2)} - \frac{(b + 3)R_4\,\xi^4}{8(b + 1)^2}\right\}.$$

This term is as far as this solution was taken, since the work involved increases very rapidly. One would expect this solution to be unreliable for $\rho \geqq 1$, but it will not be used by itself. No simplification seems likely from replacing the R's by their expressions in terms of ξ and r_1, in this case.

6. Method III. This is rather similar to the last method, involving the neglect of successive descending powers of ρ. Thus it is suitable for $\rho > 1$. In Method II a Taylor expansion about a constant was used; in this method the corresponding expansion is about a function of x. We look for a solution of (2.3) of the form

$$r_1 f(x) = mx + m_0 + m_1 x^{-1} + m_2 x^{-2} + \cdots,$$

for which it is legitimate to write

$$I_{r_1}\{r_1 f(2\rho u/r_2) \,/\, (1 + \rho)\} = I_{r_1}\{2mu/r_2\}$$
$$+ [r_1 f(2\rho u/r_2) \,/\, (1 + \rho) - 2mu/r_2]I'_{r_1}\{2mu/r_2\} + \cdots + R_n,$$

for some $n > 3$, with R_n being of order ρ^{-n}. Now, for $N \geqq n$,

$$f(2\rho u/r_2) \,/\, (1 + \rho) - 2mu/r_2) = [m_0 - 2mu/r_2]\rho^{-1}$$
$$+ [m_1 r_2/2u - (m_0 - 2mu/r_2)]\rho^{-2}$$
$$+ [m_2(r_2/2u)^2 - \{m_1 r_2/2u - (m_0 - 2mu/r_2)\}]\rho^{-3} + \cdots + T_N.$$

Thus (2.3) becomes

$$\int_0^\infty (e^{-u}u^b/b!)(\{[m_0 - 2mu/r_2]\rho^{-1} + [m_1\,r_2/2u - (m_0 - 2mu/r_2)]\rho^{-2} + \cdots\}I'_{r_1}$$
$$(2mu/r_2) + \{(1/2!)[m_0 - 2mu/r_2]\rho^{-1} + \cdots\}^2 \cdot I''_{r_1}\,(2mu/r_2) + \cdots)\,du = 0.$$

Equating to zero the coefficient of ρ^{-1} gives

$$\int_0^\infty (e^{-u}u^b/b!)(m_0 - 2mu/r_2)I'_{r_1}(2mu/r_2)\,du = 0,$$

that is, $(m/r_2)^a(1/a!b!)\int_0^\infty e^{-u(1+m/r_2)}u^{a+b}(m_0 - 2mu/r_2)\,du = 0.$

Therefore, $m_0 = 2m(a + b + 1) / (r_2 + m) = m(r_1 + r_2 - 2) / (r_2 + m)$.

Similarly, by considering the coefficients of ρ^{-2} and ρ^{-3}, we find successively

$$m_1 = \tfrac{1}{2} m(r_1 + r_2 - 2)(r_2 + m)^{-3}\{r_2(m - r_1 + 2) - 2m\},$$

$$m_2 = \tfrac{1}{6} m(r_1 + r_2 - 2)(r_2 + m)^{-5}\{2m^2(r_2 - 2)(r_2 - 4) - r_2m(3r_2^2 + 7r_1r_2 - 32r_2 - 26r_1 + 76) + r_2^2(r_1 - 2)(5r_1 + 3r_2 - 14)\}.$$

Again, this is as far as this solution was carried, due to the heavy work involved in proceeding further.

7. Final approximate solution. A *type A function* will be defined to be of the form $\phi(x) = (1 + x)^{-r}(a_{r+1}x^{r+1} + a_rx^r + a_{r-1}x^{r-1} + \cdots + a_1x + a_0)$. Now it is evident that this type of function can be put into the form of a type II or III solution, so as to agree with the first $(r + 2)$ terms in either expansion. Further solutions of the form I, II, or III (save that a finite number of terms only are considered, so that $(r + 2)$ of the calculated constants are involved) can be put into the form of type A. In this way Solution I was used to check Solutions II and III.

For a final solution a type A function is formed using Solutions II and III. Since four constants of Solution III and five of Solution II have been calculated, then $r + 2 = 4 + 5$, so that $r = 7$. The coefficients a_0, a_1, a_2, a_3, and a_4 are calculated from b_0, b_1, b_2, b_3, and b_4, and the coefficients a_8, a_7, a_6, and a_5 from m, m_0, m_1, and m_2. We put

$$a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4, \qquad a_8x^8 + a_7x^7 + a_6x^6 + a_5x^5$$

respectively equal to the corresponding terms in

$$(1 + x)^7(b_0 + b_1x + b_2x^2 + b_3x^3 + b_4x^4), \qquad (1 + x)^7(mx + m_0 + m_1x^{-1} + m_2x^{-2}).$$

Hence

$$a_0 = b_0, \qquad a_1 = b_1 + 7b_0, \qquad a_2 = b_2 + 7b_1 + 21b_0,$$

$$a_3 = b_3 + 7b_2 + 21b_1 + 35b_0, \qquad a_4 = b_4 + 7b_3 + 21b_2 + 35b_1 + 35b_0,$$

$$a_5 = m_2 + \tbinom{7}{1}m_1 + \tbinom{7}{2}m_0 + \tbinom{7}{3}m = m_2 + 7m_1 + 21m_0 + 35m,$$

$$a_6 = m_1 + \tbinom{7}{1}m_0 + \tbinom{7}{2}m = m_1 + 7m_0 + 21m,$$

$$a_7 = m_0 + \tbinom{7}{1}m = m_0 + 7\,m, \qquad a_8 = m.$$

Using these values, we take

$$r_1f(x) = (1 + x)^{-7}(a_8x^8 + a_7x^7 + \cdots + a_1x + a_0).$$

From Solutions II and III we obtain

$$m = r_1F_{r_1,r_2}(2), \qquad m_0 = m(r_2 + m)^{-1}(r_1 + r_2 - 2),$$

$$m_1 = \tfrac{1}{2}m(r_2 + m)^{-3}(r_1 + r_2 - 2)[m(r_2 - 2) - r_2(r_1 - 2)]$$
$$= \tfrac{1}{2}m_0(r_2 + m)^{-2}[m(r_2 - 2) - r_2(r_1 - 2)],$$

$$m_2 = \tfrac{1}{6}m(r_2 + m)^{-5}(r_1 + r_2 - 2)[2m^2(r_2 - 2)(r_2 - 4)$$
$$-mr_2(3r^2 + 7r_1 r_2 - 32r_2 - 26r_1 + 76) + r_2^2(r_1 - 2)(5r_1 + 3r_2 - 14)];$$

$$b_0 = \xi, \qquad b_1 = \xi, \qquad b_2 = -\tfrac{1}{2}R_2/(b + 2),$$

$$b_3 = \frac{b + 1}{(b + 2)(b + 3)}\left\{\frac{R_2^2\xi^2}{b + 1} + \frac{R_2\xi^2}{2} - \frac{R_3\xi^3}{3(b + 1)}\right\},$$

$$b_4 = \frac{(b + 1)^2}{(b + 2)(b + 3)(b + 4)}\left\{\frac{(b + 11)^4R_2R_3\xi^4}{4(b + 1)^2} - \frac{5R_2^2\xi^3}{2(b + 1)}\right.$$
$$\left. + \frac{2R_3\xi^3}{3(b + 1)} - \frac{R_2\xi^2}{2} - \frac{(b^2 + 31b + 60)R_2^3\xi^4}{8(b + 1)^2(b + 2)} - \frac{(b + 3)R_4\xi^4}{8(b + 1)^2}\right\}.$$

The solution thus derived will be called Solution IV.

8. Accuracy of solution. It can be shown that for c large compared with a,

$$\int_0^c (y^a e^{-y}/a!)\, dy = 1 - o(1).$$

Thus $r_1 f(x)/(1 + \rho)$ cannot be large compared with r_1; if it were, the left side of (2.2) would be $1 - o(1)$ and the equation would not be satisfied. So, $f(x)$ must be $O(1)$; similarly, ξ must be $O(r_1)$.

Now for Solution I to exist for large r_1 there must be at least a finite k such that $f_r(x)$ is $O(r_1^{rk})$. Since ξ is $O(r_1)$, f_0, f_1, and f_2 are of orders r_1^{0}, $r_1^{\frac{1}{2}}$, and r_1^{1}, respectively, which suggests $k = \frac{1}{2}$. That this k will suffice, or even that there exists a suitable $k \leqq 1$, has not been proved and may be only conjectured. Fortunately, it is not necessary to make any such assumption about the value.

For quick (or even any) convergence of Solution I it is necessary that $O(r_1^k) \leqq O(r_2)$ for a suitable choice of k. In practical cases r_2 is usually greater than r_1 and is often large compared with it for K positive.

It can be shown that the $f_r(x)$ of Solution I is of type A_{2r-1}, where "type A_i" will mean "of the form $(1 + x)^{-i}(a_{i+1}x^{i+1} + a_i x^i + \cdots + a_1 x + a_0)$." If Solution I were developed as far as $f_4(x)$, it would yield a type A_7 function, which could be compared with the A_7 function Solution IV. The former of these two type A_7 functions is correct to the order r_2^{-4}. Thus when it is expanded in ascending or descending powers of x, the resulting coefficients are correct to the order r_2^{-4}, and thus differ from the exact ones obtained from Methods II and III, respectively, by terms of the order r_2^{-5}.

Consequently it is readily seen that the type A_7 form of Solution I and that of Solution IV differ only by terms of the order r_2^{-5}, so that Solution IV is correct to the order r_2^{-4}. Using the consequences of the above discussion, we see that Solution IV is correct to the order $(r_2/r_1^k)^{-4}$.

Further, if Solution IV is put in place of $f(x)$ in (2.2), the error involved for ρ small will be of the order ρ^5. Similarly the error involved in using Solution IV for ρ large is of the order ρ^{-4}. These statements apply whether or not there exists an $f(x)$ exactly satisfying this equation.

For convenience of calculation one could, of course, use fewer leading terms of Solutions II and III to form a less accurate Solution IV. The error involved in substituting this Solution IV into (2.2) may be rather less than the error in the approximate $f(x)$, particularly if the upper limit of the inner integrand is sufficiently large or small, when the rate of change of the inner integral with respect to the upper limit will be negligible.

However, when tables have been prepared using the solution given in this paper, there will be no need to use a less accurate approximation to save labour.

The solution given in Section 7, say $f_{IV}(x)$, has been calculated for $r_1 = 8$, $r_2 = 50$, $\alpha = .975$, and a series of x values. The left side of (2.2) was then calculated for $\rho = 1$. One would not expect this to give a particularly accurate value to a function correct to the orders ρ^5 for ρ small and ρ^{-4} for ρ large. Further $r_2/r_1 = 6.25$, which is not very large. Thus one would expect the majority of practical cases to be more favourable than that chosen. By numerical integration, the value of the left side of (2.2) was found correct to five significant figures as .97492, a satisfactory approximation to α.

9. Obtaining and using the confidence limit. If f_{IV} is of suitable form, a suitable approximate confidence limit for K will be given by solving

$$M_1/\mathrm{K} = f_{IV}(M_2/\mathrm{K}),$$

where f_{IV} is the function given by Solution IV. In view of the complicated form of $f_{IV}(x)$, a numerical method of solution evidently will be necessary. Since $x = M_2/\mathrm{K}$ and $y = M_1/\mathrm{K}$, the ratio y/x is M_1/M_2, which has an observed value. Thus the confidence limit, K_α, is given by the intersection (x_0, y_0) of the curve $y = f_{IV}(x)$ with the line $y = (M_1/M_2)x$, since

$$\mathrm{K}_\alpha = M_1/y_0 = M_2/x_0 .$$

This gives a lower limit such that $\Pr(\mathrm{K}_\alpha \leq \mathrm{K}) = \alpha$.

Certain questions arise immediately:

(*i*). How can one be sure that there will be only one point of intersection?

(*ii*). How can one be sure that there will be any point of intersection?

(*iii*). The previous results depend on the assumption that $\mathrm{K} > 0$. What modifications are required for the case where the sign of K is not known?

These matters will be considered in turn.

(*i*). The complicated expression for f_{IV} makes uniqueness of intersection difficult to prove. A sufficient condition for having no more than one point of intersection (since the asymptote to the curve and curve itself intersect the y-axis in positive values of y) is that the slope of the curve $y = f_{IV}(x)$ should be a monotonically increasing or decreasing function of x. The asymptote to the curve is given by the first two terms of Solution III, the second of which (that independent of x) is positive. This condition is satisfied in the particular example mentioned at the end of the previous section, in which the slope is monotonically increasing. Further, the condition is satisfied when r_2 is sufficiently large com-

pared with r_1; also, the slope of the curve is increasing when α is chosen to correspond to the lower limit, and decreasing in the upper limit case, provided the two appropriate choices of α are both reasonably different from 50 per cent. Hence, at least for r_2 sufficiently large compared with r_1, if not more generally (as the author conjectures), there is no more than one confidence limit corresponding to one value of α.

(ii). When the slope is monotonically increasing or decreasing, evidently there will be no intersection of the line with the curve, unless M_1/M_2 is greater than or equal to the asymptotic slope of the curve, which equals $F_{r_1,r_2}(\alpha)$. Now M_1/M_2 is distributed as

$$(\sigma_1^2/\sigma_2^2)F_{r_1,r_2} = [(1 + \rho)/\rho]\, F_{r_1,r_2} .$$

Thus the probability of nonintersection is the probability of an F_{r_1,r_2} variate not exceeding $\rho F_{r_1,r_2}(\alpha)/(1 + \rho)$. This probability is α when $\rho = \infty$, and decreases with decreasing ρ to zero at $\rho = 0$. Since M_1 and M_2 are positive, it is evident that an intersection in the first quadrant leads to a positive K_α. Further, it can easily be shown that, as $M_1/M_2 \to F_{r_1,r_2}(\alpha)$ from above, $\mathrm{K}_\alpha \to 0$ from above.

At this stage it is convenient to consider *(iii)* along with *(ii)*. Now all the previous investigation of a solution providing a confidence limit for K could have been treated in exactly the same way with M_1 and M_2 and all related quantities interchanged. In this way an approximation to a function h, such that

$$\Pr\{M_2/(-\mathrm{K}) \leqq h\,[M_1/(-K)]\} = \alpha$$

independently of a nuisance parameter

$$\rho' = -\mathrm{K}/\sigma_1^2 = -\mathrm{K}/(\mathrm{K} + \sigma_2^2) = -1/(1 + \rho^{-1}) = -\rho/(1 + \rho)$$

would have been obtained. An approximate h_{IV} would have been derived equal to f_{IV} with interchanged r_1 and r_2. A variation of ρ' between 0 and ∞ corresponds to one of ρ between 0 and -1, and is appropriate for K negative. The interchanged ranges are appropriate for K positive. The accuracy of h_{IV} with respect to ρ' would be the same as that of f_{IV} with respect to ρ as far as neglected orders are concerned. However, if r_2 is large compared with r_1, favouring the accuracy of f_{IV}, then the accuracy of h_{IV} would not be so favoured; the reverse is true for r_1 large compared with r_2.

It will be seen to be satisfactory to use the curve $y = f_{IV}(x)$ in the first quadrant together with a second curve, $-x = h_{IV}(-y)$, in the third to provide a confidence limit. However, if the coefficient α is chosen for the first curve, then $1 - \alpha$ will be taken for the second. The reasons for this are explained in the following paragraphs.

Since $F_{r_1,r_2}(\alpha)\, F_{r_2,r_1}(1 - \alpha) = 1$, the asymptote to the first curve in the first quadrant will be parallel to the asymptote to the second curve in the third. Thus a straight line through the origin will intercept the first curve if its slope is less than $F_{r_1,r_2}(\alpha)$, while if its slope is exactly equal to this value, it intersects both curves at infinity.

The set of points S, consisting of those points of the first quadrant lying below the first curve and those of the third quadrant lying above the second curve, will be used to obtain a confidence limit for K. If $K > 0$, the probability density function is zero outside the first quadrant and the probability of M_1 and M_2 being such that (x, y) lies beneath the first curve is α. If $K < 0$, the probability density function is zero outside the third quadrant, and the probability of (x, y) lying above the second curve is α.

Thus, whether K is greater or less than zero, the probability is α that M_1 and M_2 are such that (x, y) lies in S. Just as an intersection of the first curve with the line $y = (M_1/M_2)x$ gives a positive value of K_α, an intersection of the second curve with this line gives a negative value. Further, it can be shown easily that $K_\alpha \to 0$ from below as $M_1/M_2 \to F_{r_1,r_2}(\alpha)$ from below, or as $M_2/M_1 \to F_{r_2,r_1}(1 - \alpha)$ from above.

Thus the two curves together provide a lower confidence limit which falls below K with probability α. Evidently they provide equivalently an upper limit with coefficient $1 - \alpha$. Accordingly, two suitable values of α are selected, one for each limit, and an interval is obtained. The values .025 and .975, giving an interval coefficient of .95, are frequently used in practice. The complicated form of K_α does not lend itself to an examination of which pair of values of α having a given difference (confidence coefficient) yield the shortest interval.

Incidentally, the curves obtained by imaging radially the two curves through the origin into the opposite quadrants can be shown easily to form the curved parts of the boundary of an alternative set of points which yields a confidence limit. However, it is usual to have $K > 0$ and $r_2 > r_1$, and one would prefer the positive confidence limit to be more accurate. To ensure this, the two curves should be used as discussed above.

Under *(iii)*, it remains to be decided whether or not the confidence coefficient is affected by using only that part of the confidence interval which has the same sign as K, if this sign is known. Consider the lower limit with coefficient α_1, when K is known to be positive

$$\begin{aligned}\Pr\{\max(0, K_{\alpha_1}) \leqq K \mid K > 0\} &= 1 - \Pr\{\max(0, K_{\alpha_1}) \geqq K \mid K > 0\}\\ &= 1 - \Pr\{K_{\alpha_1} \geqq K \mid K > 0\}\\ &= 1 - (1 - \alpha_1) = \alpha_1 = \Pr\{K_{\alpha_1} \leqq K\}.\end{aligned}$$

A similar discussion applies when K is known to be negative, also for the upper limit when K has known sign. Hence the natural procedure does not distort the confidence coefficient.

In the Introduction, we discussed the use of this confidence interval, or a single limit, for testing a hypothetical value of K. When K is known to be greater than or equal to zero and the hypothesis to be tested is $K = 0$, the use of the upper limit alone is more appropriate. In this case the hypothesis is rejected if $M_1/M_2 > F_{r_1,r_2}(\alpha)$, which is the usual test of the analysis of variance.

However, throughout this paper there is one possibility which has not been

discussed—nor is it obvious how it could be, considering the technique that has been used. This is the case of $K = 0$, when the above derivation of a confidence interval would be completely invalid. That is not to say that the interval does not apply in this case—(the author conjectures that it does), but the point is just not proved either way, although it is true that $\Pr[y/x \leqq f(x)/x] = \alpha$ in the limit as $K \to 0$. Thus the confidence interval carries with it the perhaps unnecessary proviso that K is not zero.

10. Tabulation. For the practical use of the two curves, discussed in the preceding section, to obtain a confidence limit, the following procedure seems the most satisfactory. For each selected value of α, the values of $f_{IV}(x)/x$ are tabulated for different values of r_1, r_2, and x. It might then be advisable to retabulate, so that for each set of values of r_1, r_2, and $f_{IV}(x)/x$ a value of x, or of $f_{IV}(x)$, is tabulated; otherwise use of the table would require inverse interpolation. To use the hypothetical table,

If $M_1/M_2 \geqq F_{r_1,r_2}(\alpha)$, it is set equal to $f_{IV}(x)/x$ and, by direct interpolation, the appropriate value of $x = M_2/K_\alpha$ is obtained and since M_2 is known, K_α can then be derived;

if $M_1/M_2 = F_{r_1,r_2}(\alpha)$, then $K_\alpha = 0$;

if $M_1/M_2 < F_{r_1,r_2}(\alpha)$, then r_2, r_1, and $(1 - \alpha)$ are used as new values of r_1, r_2, and α, respectively, in the first procedure.

The expression for $f_{IV}(x)$ is very complicated, and the tabulation discussed above, for a suitable selection of values of r_1, r_2, and x, would require tens of thousands of cells. Accordingly the table has not been constructed for inclusion in this paper, and that task remains.

Acknowledgements: The author wishes to express his gratitude to Mr. R. L. Plackett, who suggested the problem to him and supervised his research. Thanks are also due to Dr. B. L. Welch, who made helpful criticisms of an earlier draft of this paper.

REFERENCES

[1]. A. A. Aspin, "An examination and further development of a formula arising in the problem of comparing two mean values," *Biometrika*, Vol. 35 (1948), pp. 88–96.

[2] A. A. Aspin, "Tables for use in comparisons whose accuracy involves two variances, separately estimated," *Biometrika*, Vol. 36 (1949), pp. 290–293.

[3] I. Bross, "Fiducial intervals for variance components," *Biometrics*, Vol. 6 (1950), pp. 136–144.

[4] R. E. Comstock and H. F. Robinson, "The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance," *Biometrics*, Vol. 4 (1948), pp. 254–266.

[5] S. L. Crump, "The estimation of variance components in analysis of variance," *Biometrics*, Vol. 2 (1946), pp. 7–11.

[6] S. L. Crump and A. W. Nordskog, "Systematic and random sampling for estimated egg production in poultry," *Biometrics*, Vol. 4 (1948), pp. 223–233.

[7] S. L. Crump, "The present status of variance component analysis," *Biometrics*, Vol. 7 (1951), pp. 1–16.

[8] H. E. Daniels, "The estimation of components of variance," *J. Roy. Stat. Soc. Suppl.*, Vol. 6 (1939), pp. 186–197.

[9] R. A. Fisher, "The fiducial argument in statistical inference," *Ann. Eugenics*, Vol. 6 (1935), pp. 391–398.

[10] S. Marcuse, "Optimum allocation and variance components in nested sampling with an application to chemical analysis," *Biometrics*, Vol. 5 (1949), pp. 189–206.

[11] K. Pearson, "On the application of the double Bessel function, $K_{t_1 \cdot t_2}(x)$, to statistical problems," *Biometrika*, Vol. 25 (1933), pp. 158–178.

[12] L. H. C. Tippett, *The Methods of Statistics*, Williams and Norgate, London, 1931.

[13] B. L. Welch, "The generalization of 'Student's' problem when several different population variances are involved," *Biometrika*, Vol. 34 (1947), pp. 28–35.

[14] B. L. Welch, "On the Studentization of several variances," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 118–122.

[15] F. Yates, "Principles governing the amount of experimentation in developmental work," *Nature*, Vol. 170 (1952), pp. 138–140.

A SIMPLE SEQUENTIAL PROCEDURE FOR TESTING STATISTICAL HYPOTHESES[1]

By Chia Kuei Tsao

Wayne University

Summary. In this paper a simple sequential test is suggested. Distribution of the sample size, its moment generating function, the power function of the test, and the ASN (average sample number) function are obtained. The determination of the set of relative optimum zones for making decisoins is shown to be unique. The existence of a class of sets of absolute optimum zones is proved. The suggested test is shown to be consistent. Some possible applications are discussed and a few numerical efficiencies are calculated.

1. Introduction. Let $\{f(x)\}$ be the class of all continuous pdf's (probability distribution functions) defined over a space S. Let random observations be drawn successively from a population having an unknown continuous pdf $f(x)$. Let the simple hypothesis $H_0:f(x) = f_0(x)$ be tested against a certain alternative or a certain class of alternatives. We shall propose a simple sequentual test procedure and be concerned with the investigation of the properties of the test.

To test the null hypothesis $H_0:f(x) = f_0(x)$, we divide S into three mutually exclusive sets (zones):

S_1 is the zone of preference for acceptance;

S_2 is the zone of indifference;

S_3 is the zone of preference for rejection.

Random observations are drawn successively. At each stage, the number of observations falling in each of the three zones will be counted. Let m_i be the number of observations falling in the zone S_i for $i = 1, 2, 3$ at the mth stage (i.e., after the mth observation has been drawn). Let a and r be two predetermined positive integers. Continue to draw observations as long as $m_1 < a$ and $m_3 < r$. The experiment is discontinued as soon as either $m_1 = a$ or $m_3 = r$. The null hypothesis is accepted if $m_1 = a$, and rejected if $m_3 = r$.

For simplicity, we shall restrict S to be n-dimensional Euclidean space (or a subspace of it) and assume, of course, that the pdf $f(x)$ is continuous in S. However, most of the theorems given in this paper can be extended to more general cases with slight modifications.

2. Fundamental lemma. The principal aim of this section is to prove a lemma which was used for obtaining the moment generating function of the sample size and the power function of the test.

Suppose m, p, and q are positive integers and B, C, and D are positive real

Received: 4/27/53, revised 3/22/54.

[1] This work was supported in part by the Office of Naval Research.

numbers. Let

$$
\begin{aligned}
h_m(p, q, B, C, D) &= \sum_{x=0}^{p-1} \frac{(m-1)!}{(q-1)!x!(m-q-x)!} B^q C^x D^m, \\
h(p, q, B, C, D) &= \sum_{m=q}^{\infty} h_m(p, q, B, C, D).
\end{aligned} \tag{2.1}
$$

Then we have the following

LEMMA 1. *If $D < 1$ and $D + CD < 1$, then*

$$
\begin{aligned}
&h(p, q, B, C, D) \\
&= \left(\frac{BD}{1-D-CD}\right)^q \left(1 - \int_0^{CD/(1-D)} \frac{(p+q-1)!}{(p-1)!(q-1)!} z^{p-1}(1-z)^{q-1}\, dz\right).
\end{aligned} \tag{2.2}
$$

PROOF. From definition (2.1) we have

$$
\begin{aligned}
h(p, q, B, C, D) &= \sum_{m=q}^{\infty} \sum_{x=0}^{p-1} \frac{(m-1)!}{(q-1)!x!(m-q-x)!} B^q C^x D^m \\
&= \sum_{x=0}^{p-1} \frac{B^q C^x D^{q+x}}{(q-1)!x!} \sum_{m=q+x}^{\infty} \frac{(m-1)!}{(m-q-x)!} D^{m-q-x} \\
&= \sum_{x=0}^{p-1} \frac{B^q C^x D^{q+x}}{(q-1)!x!} (q+x-1)!(1-D)^{-(q+x)} \\
&= \sum_{x=0}^{p-1} \frac{(q+x-1)!}{(q-1)!x!} \left(\frac{BD}{1-D}\right)^q \left(\frac{CD}{1-D}\right)^x \\
&= \left(\frac{BD}{1-D-CD}\right)^q \sum_{x=0}^{p-1} \binom{q+x-1}{x} \\
&\qquad \cdot \left(\frac{1-D-CD}{1-D}\right)^q \left(\frac{CD}{1-D}\right)^x \\
&= \left(\frac{BD}{1-D-CD}\right)^q \\
&\qquad \cdot \left(1 - \int_0^{CD/(1-D)} \frac{(p+q-1)!}{(p-1)!(q-1)!} z^{p-1}(1-z)^{q-1}\, dz\right).
\end{aligned} \tag{2.3}
$$

Thus, Lemma 1 is proved.

3. Distribution of the sample size and its moment generating function. The distribution of the sample size, (a, r, S_1, S_2, S_3) being chosen, depends upon the true underlying distribution, $f(x)$, being tested. In this section, we derive the pdf $g_f(m;\ a, r, S_1, S_3)$ of the sample size m and its mgf (moment generating function) $M_f(t;\ a, r, S_1, S_3)$ under the assumption that $f(x)$ is the true underlying distribution and the set of parameters (a, r, S_1, S_2, S_3) is predetermined.

Throughout this paper, we shall denote by A, I, and R the following three

quantities:

$$A = \int_{S_1} f(x)\,dx, \qquad I = \int_{S_2} f(x)\,dx, \qquad R = \int_{S_3} f(x)\,dx. \tag{3.1}$$

We shall denote these quantities by

(a) A_i, I_i, and R_i, if $f(x)$ is replaced by $f_i(x)$ for $i = 0$ or 1;

(b) A', I', and R', if (S_1, S_2, S_3) is replaced by (S_1', S_2', S_3');

(c) A_i', I_i', and R_i', if $f(x)$ is replaced by $f_i(x)$ for $i = 0$ or 1, and (S_1, S_2, S_3) by (S_1', S_2', S_3').

With the definitions (3.1), it is easily seen that the pdf is given by

$$\begin{aligned} g_f(m;\ a, r, S_1, S_3) = {} & \sum_{x=0}^{a-1} \frac{(m-1)!}{(r-1)!x!(m-r-x)!} R^r A^x I^{m-r-x} \\ & + \sum_{x=0}^{r-1} \frac{(m-1)!}{(a-1)!x!(m-a-x)!} A^a R^x I^{m-a-x}. \end{aligned} \tag{3.2}$$

Therefore the mgf is given by

$$\begin{aligned} M_f(t;\ a, r, S_1, S_3) = {} & \sum_{m=r}^{\infty} \sum_{x=0}^{a-1} \frac{(m-1)!}{(r-1)!x!(m-r-x)!} R^r A^x I^{m-r-x} e^{mt} \\ & + \sum_{m=a}^{\infty} \sum_{x=0}^{r-1} \frac{(m-1)!}{(a-1)!x!(m-a-x)!} A^a R^x I^{m-a-x} e^{mt}. \end{aligned} \tag{3.3}$$

This can be written as

$$\begin{aligned} M_f(t;\ a, r, S_1, S_3) = {} & \sum_{m=r}^{\infty} \sum_{x=0}^{a-1} \frac{(m-1)!}{(r-1)!x!(m-r-x)!} \left(\frac{R}{I}\right)^r \left(\frac{A}{I}\right)^x (Ie^t)^m \\ & + \sum_{m=a}^{\infty} \sum_{x=0}^{r-1} \frac{(m-1)!}{(a-1)!x!(m-a-x)!} \left(\frac{A}{I}\right)^a \left(\frac{R}{I}\right)^x (Ie^t)^m \\ = {} & h(a, r, R/I, A/I, Ie^t) + h(r, a, A/I, R/I, Ie^t). \end{aligned} \tag{3.4}$$

Thus, by Lemma 1, the mgf can be written as

$$\begin{aligned} M_f(t;\ a, r, S_1, S_3) = {} & \left(\frac{Re^t}{1-(1-R)e^t}\right)^r \\ & \cdot \left(1 - \int_0^{Ae^t/(1-Ie^t)} \frac{(a+r-1)!}{(a-1)!(r-1)!} z^{a-1}(1-z)^{r-1}\,dz\right) \\ & + \left(\frac{Ae^t}{1-(1-A)e^t}\right)^a \left(1 - \int_0^{Re^t/(1-Ie^t)} \frac{(r+a-1)!}{(r-1)!(a-1)!} z^{r-1}(1-z)^{a-1}\,dz\right). \end{aligned} \tag{3.5}$$

4. The power and ASN functions. Suppose the set of parameters (a, r, S_1, S_2, S_3) is predetermined. Then, it is easily seen that for any alternative $f(x)$, the power function is given by

$$\varphi(f;\ a, r, S_1, S_3) = \sum_{m=r}^{\infty} \sum_{x=0}^{a-1} \frac{(m-1)!}{(r-1)!x!(m-r-x)!} R^r A^x I^{m-r-x}. \tag{4.1}$$

By Lemma 1, this can be written as

$$
\begin{aligned}
\varphi(f;\quad a, r, S_1, S_3) &= h(a, r, R/I, A/I, I), \\
&= 1 - \int_0^{1-\beta} \frac{(a + r - 1)!}{(a - 1)!(r - 1)!} z^{a-1}(1 - z)^{r-1}\, dz, \qquad (4.2)\\
&= \int_0^{\beta} \frac{(a + r - 1)!}{(a - 1)!(r - 1)!} z^{r-1}(1 - z)^{a-1}\, dz,
\end{aligned}
$$

where $\beta = \beta(f;\quad S_1, S_3) = 1/(1 + A/R)$.

By the use of the mgf (3.5), it is easily verified that the average sample number (ASN) is given by

$$
\begin{aligned}
\mu(f;\quad a, r, S_1, S_3) &= \frac{r}{R}\left[\varphi(f;\quad a, r, S_1, S_3) - \binom{r + a - 1}{r}\beta^r(1 - \beta)^a\right] \qquad (4.3)\\
&\quad + \frac{a}{A}\left[1 - \varphi(f;\quad a, r, S_1, S_3) - \binom{r + a - 1}{a}\beta^r(1 - \beta)^a\right].
\end{aligned}
$$

This can also be shown to be

$$
\begin{aligned}
\mu(f;\quad a, r, S_1, S_3) &= \frac{r}{R}\left[1 - \int_0^{1-\beta} \frac{(a + r)!}{(a - 1)!r!} z^{a-1}(1 - z)^r\, dz\right] \qquad (4.4)\\
&\quad + \frac{a}{A}\left[1 - \int_0^{\beta} \frac{(r + a)!}{(r - 1)!a!} z^{r-1}(1 - z)^a\, dz\right].
\end{aligned}
$$

5. Optimum zones S_1, S_2, S_3. In testing the null hypothesis $f_0(x)$ against an alternative hypothesis $f_1(x)$, all the four quantities

$$
\varphi(f_0;\ a, r, S_1, S_3), \qquad \varphi(f_1;\ a, r, S_1, S_3),
$$

$$
\mu(f_0;\ a, r, S_1, S_3), \qquad \mu(f_1;\ a, r, S_1, S_3)
$$

are functions of the four parameters a, r, S_1, and S_3. Accordingly, there may be many ways of defining the optimum zones. However, in choosing a definition, we should take into consideration the following three problems: (a) the definition itself should be reasonable from the point of view of the statistician; (b) it must be realizable, that is, the optimum zones must exist; and (c) it can be put in a form suitable for applications.

Furthermore, if the pair of positive integers (a, r) is preassigned, a set of optimum zones should be such that it is optimum (in some sense) among all possible sets (S_1, S_2, S_3). If the pair (a, r) is to be determined by the experimenter, then a set should be so chosen that it has certain optimum properties in the whole parameter space $\{(a, r, S_1, S_2, S_3)\}$, that is, it is optimum for all possible choices of pairs (a, r) and all possible choices of sets (S_1, S_2, S_3). In the following, we give two definitions, one for a fixed pair (a, r) and the other for the general case. However, the determination of the optimum zones for the general case is so difficult that we shall just prove their existence.

For any given set (α, φ, a, r), where $0 < \alpha < \varphi < 1$, we shall denote by $\Omega_{\alpha,\varphi,a,r}$ the class of all possible sets of the three zones (S_1, S_2, S_3) which satisfy the following two conditions:

$$\varphi(f_0;\ a, r, S_1, S_3) = \alpha, \qquad \varphi(f_1;\ a, r, S_1, S_3) = \varphi. \tag{5.1}$$

Here, we have assumed that the class $\Omega_{\alpha,\varphi,a,r}$ is nonempty. A proof of the existence of such a class under certain general conditions will be given in Section 6.

A test is said to have the *strength* (α, φ), if its power function satisfies the two conditions (5.1). Thus, every test based on a set of $\Omega_{\alpha,\varphi,a,r}$ has the strength (α, φ).

Definition I. A set (S_1, S_2, S_3) of $\Omega_{\alpha,\varphi,a,r}$ is said to be *relatively optimum* with respect to (a, r), if the inequalities

$$\begin{aligned} &\mu(f_0;\ a, r, S_1, S_3) \leqq \mu(f_0;\ a, r, S_1', S_3'), \\ &\qquad\qquad \mu(f_1;\ a, r, S_1, S_3) \leqq \mu(f_1;\ a, r, S_1', S_3') \end{aligned} \tag{5.2}$$

hold for all sets $(S_1', S_2', S_3') \,\varepsilon\, \Omega_{\alpha,\varphi,a,r}$. The three zones of a relative optimum set are called *relative optimum zones.*

To determine the relative optimum zones, we need first to prove the following two lemmas.

Lemma 2. *For fixed a and r, the* ASN *function $\mu(f;\ a, r, S_1, S_3)$ decreases as either A or R increases.*

Proof. Taking the partial derivatives of the ASN function (4.4) with respect to A and R, we obtain

$$\frac{\partial\mu}{\partial R} = -\frac{r}{R^2}\left[1 - \int_0^{1-\beta} \frac{(a+r)!}{(a-1)!r!} z^{a-1}(1-z)^r\, dz\right], \tag{5.3}$$

$$\frac{\partial\mu}{\partial A} = -\frac{a}{A^2}\left[1 - \int_0^{\beta} \frac{(r+a)!}{(r-a)!a!} z^{r-1}(1-z)^a\, dz\right]. \tag{5.4}$$

Since (5.3) and (5.4) are always negative, Lemma 2 is proved.

Lemma 3. *Suppose* (a) *$f_0(x)$ and $f_1(x)$ are continuous,* (b) *for every real number c, the probability measure of the set $\{x;\ f_1(x)/f_0(x) = c\}$ under either hypothesis is zero, and* (c) *the set (S_1, S_2, S_3) defined by*

$$S_1 = \{x;\ f_1(x)/f_0(x) \leqq k_0\}, \tag{5.5}$$

$$S_2 = \{x;\ k_0 \leqq f_1(x)/f_0(x) \leqq k_1\}, \tag{5.6}$$

$$S_3 = \{x;\ k_1 \leqq f_1(x)/f_0(x)\}, \tag{5.7}$$

where $k_0 \leqq k_1$ are two constants, belongs to $\Omega_{\alpha,\varphi,a,r}$.

Then, for any set (S_1', S_2', S_3') in $\Omega_{\alpha,\varphi,a,r}$, we have

$$A_0' \leqq A_0, \qquad R_0' \leqq R_0, \qquad A_1' \leqq A_1, \qquad R_1' \leqq R_1. \tag{5.8}$$

Proof. In order to prove Lemma 3, it is sufficient to prove (i) if $A_0' \leqq A_0$,

then $R_0' \leqq R_0$, $A_1' \leqq A_1$ and $R_1' \leqq R_1$, and (ii) under the given assumptions, the inequality $A_0' \leqq A_0$ holds.

First, assume $A_0' \leqq A_0$. Since both (S_1, S_2, S_3) and (S_1', S_2', S_3') are in $\Omega_{\alpha,\varphi,a,r}$, then, by (4.2) and (5.1), we have

$$(5.9) \qquad \text{(a)} \quad A_0'R_0' = A_0/R_0, \qquad \text{(b)} \quad A_1'R_1' = A_1R_1.$$

By (5.9a), $A_0' \leqq A_0$ implies $R_0' \leqq R_0$. By (5.7), $R_0' \leqq R_0$ implies $R_1' \leqq R_1$ (using an argument of the Neyman-Pearson type). Finally, by (5.9b), $R_1' \leqq R_1$ implies $A_1' \leqq A_1$, proving (i).

Next, assume $A_0' > A_0$. Then, by (5.9a), there exists a positive number δ such that

$$(5.10) \qquad A_0' = A_0 + \delta A_0, \qquad R_0' = R_0 + \delta R_0.$$

Therefore, by (5.5), (5.6), (5.7), we must have

$$(5.11) \qquad A_1' > A_1 + \delta A_1, \qquad R_1' < R_1 + \delta R_1,$$

which imply that $A_1'/R_1' > A_1/R_1$. Consequently, we obtain

$$(5.12) \qquad \varphi(f_1;\ a, r, S_1', S_3') < \varphi(f_1;\ a, r, S_1, S_3).$$

This contradicts the assumption that both (S_1, S_2, S_3) and (S_1', S_2', S_3') are members of $\Omega_{\alpha,\varphi,a,r}$. Hence, the inequality $A_0' \leqq A_0$ must hold, proving (ii), which completes the proof of Lemma 3.

THEOREM 1. *Under the conditions given in Lemma* 3, *the set of the relative optimum zones* (S_1, S_2, S_3) *with respect to* (a, r) *for testing the simple hypothesis* $f_0(x)$ *against the alternative hypothesis* $f_1(x)$ *with strength* (α, φ) *is the set determined by* (5.5), (5.6), *and* (5.7)

Theorem 1 follows from Lemmas 2 and 3.

From Theorem 1, it is seen that, for each (α, φ, a, r), the set of relative optimum zones (S_1, S_2, S_3), when it exists, is uniquely determined. In the following, we shall assume that, for every (α, φ, a, r), the set of the relative optimum zones exists. We shall denote by $\Omega_{\alpha,\varphi}$ the class of all possible sets of the three zones such that the corresponding tests will all have strength (α, φ) for testing $f_0(x)$ against $f_1(x)$, that is, $\Omega_{\alpha,\varphi} = \bigcup_{a,r} \Omega_{\alpha,\varphi,a,r}$. To distinguish the sets in $\Omega_{\alpha,\varphi}$ from the sets in $\Omega_{\alpha,\varphi,a,r}$ for some fixed (a, r), we shall write (a, r, S_1, S_2, S_3) as the general set in $\Omega_{\alpha,\varphi}$. We shall also denote by $\Omega_{\alpha,\varphi,0}$ the class of all sets of the relative optimum zones in $\Omega_{\alpha,\varphi}$, that is, all sets (a, r, S_1, S_2, S_3), where, for each pair (a, r), the set (S_1, S_2, S_3) is the set of relative optimum zones with respect to (a, r).

A set (a, r, S_1, S_2, S_3) of $\Omega_{\alpha,\varphi}$ is said to be *comparable* with another set $(a', r', S_1', S_2', S_3')$ of $\Omega_{\alpha,\varphi}$ if either the two inequalities,

$$(5.13) \qquad \begin{aligned} &\mu(f_0;\ a, r, S_1, S_3) \leqq \mu(f_0;\ a', r', S_1', S_3'), \\ &\mu(f_1;\ a, r, S_1, S_3) \leqq \mu(f_1;\ a', r', S_1', S_3'), \end{aligned}$$

hold simultaneously, or the two inequalities,

$$(5.14) \qquad \begin{aligned} &\mu(f_0;\quad a, r, S_1, S_3) \geqq \mu(f_0;\quad a', r', S_1', S_3'), \\ &\qquad\qquad \mu(f_1;\quad a, r, S_1, S_3) \geqq \mu(f_1;\quad a', r', S_1', S_3'), \end{aligned}$$

hold simultaneously. Otherwise, they are said to be *noncomparable*. Two comparable sets are said to be *equivalent*, if all four inequalities in (5.13) and (5.14) hold simultaneously.

LEMMA 4. *Given any set* (a, r, S_1, S_2, S_3) *in* $\Omega_{\alpha,\varphi}$, *there is a set* $(a', r', S_1', S_2', S_3')$ *in* $\Omega_{\alpha,\varphi,0}$ *such that the two inequalities* (5.14) *hold simultaneously.*

The proof is trivial, since we can always choose $a' = a$ and $r' = r$.

LEMMA 5. *For any set* (a, r, S_1, S_2, S_3) *in* $\Omega_{\alpha,\varphi}$, *the number of sets* $(a', r', S_1', S_2', S_3')$ *in* $\Omega_{\alpha,\varphi,0}$ *satisfying the two inequalities* (5.14) *is finite.*

PROOF. For any set $(a', r', S_1', S_2', S_3')$ of $\Omega_{\alpha,\varphi,0}$ (or of $\Omega_{\alpha,\varphi}$ in general), the following two inequalities,

$$(5.15) \qquad \begin{aligned} &\mu(f_0;\quad a', r', S_1', S_3') \geqq r'\alpha + a'(1-\alpha), \\ &\qquad\qquad \mu(f_1;\quad a', r', S_1', S_3') \geqq r'\varphi + a'(1-\varphi), \end{aligned}$$

must hold. But, for any set (a, r, S_1, S_2, S_3) in $\Omega_{\alpha,\varphi}$, the two quantities $\mu(f_0;\quad a, r, S_1, S_3)$ and $\mu(f_1;\quad a, r, S_1, S_3)$ are finite. Thus, Lemma 5 follows from the uniqueness of the set of relative optimum zones for each (a', r').

From Lemmas 4 and 5, it is obvious that, for each set (a, r, S_1, S_2, S_3) in $\Omega_{\alpha,\varphi}$, there exists a comparable set $(a^*, r^*, S_1^*, S_2^*, S_3^*)$ in $\Omega_{\alpha,\varphi,0}$ such that the following two inequalities,

$$(5.16) \qquad \begin{aligned} &\mu(f_0;\quad a^*, r^*, S_1^*, S_3^*) \leqq \mu(f_0;\quad a, r, S_1, S_3), \\ &\qquad\qquad \mu(f_1;\quad a^*, r^*, S_1^*, S_3^*) \leqq \mu(f_1;\quad a, r, S_1, S_3), \end{aligned}$$

hold for all sets (a, r, S_1, S_2, S_3) in $\Omega_{\alpha,\varphi}$ which are comparable with $(a^*, r^*, S_1^*, S_2^*, S_3^*)$. Denoting by $\Omega^*_{\alpha,\varphi,0}$ the class of all such sets $(a^*, r^*, S_1^*, S_2^*, S_3^*)$ in $\Omega_{\alpha,\varphi}$, we may conclude:

THEOREM 2. *The class* $\Omega^*_{\alpha,\varphi,0}$ *is a subclass of* $\Omega_{\alpha,\varphi,0}$. *Two distinct sets in* $\Omega^*_{\alpha,\varphi,0}$ *are either equivalent or noncomparable.*

The class $\Omega^*_{\alpha,\varphi,0}$ may be called the class of sets of the *absolute optimum zones*. Since there may be many sets of the absolute optimum zones and the determination of any such set is difficult, we shall assume, throughout the remaining part of this paper, that a and r are preassigned and the three zones are chosen according to (5.5), (5.6), and (5.7). We shall also denote by $\varphi(f)$ and $\mu(f)$ the power and the ASN functions of the test if the three zones are so chosen.

We have seen that, for each (α, φ, a, r), the set of the relative optimum zones (S_1, S_2, S_3), when it exists, is uniquely determined. On the other hand, it is easily seen that, for each (α, a, r), there are an infinite number of sets of the relative optimum zones (S_1, S_2, S_3). We shall denote by $\Omega_{\alpha,a,r,0}$ the class of all such sets of the relative optimum zones (S_1, S_2, S_3), that is

$$\Omega_{\alpha,a,r,0} = \{\Omega_{\alpha,\varphi,a,r} \cap \Omega_{\alpha,\varphi,0};\qquad \alpha < \varphi < 1\}.$$

The test based on a preassigned (α, a, r) and a set (S_1, S_2, S_3) will be called the R. O. (relatively optimum) test with respect to (a, r) for fixed level of significance α, or simply the R. O. test, if $(S_1, S_2, S_3) \varepsilon \Omega_{\alpha,a,r,0}$.

6. Consistency and existence of $\Omega_{\alpha,\varphi,a,r}$. In Section 5, we assumed that, for any given set (α, φ, a, r), the class $\Omega_{\alpha,\varphi,a,r}$ is nonempty. This assumption is valid only when $f_0(x)$ and $f_1(x)$ satisfy certain general conditions. On the other hand, the consistency of the R. O. test depends on the existence of such classes.

Suppose a and r are presassigned positive integers. Suppose the hypothesis $f_0(x)$ is to be tested against the alternative hypothesis $f_1(x)$. Again, we shall assume that $f_0(x)$ and $f_1(x)$ are continuous, and, for every real number c, the probability measure of the set $\{x;\ f_1(x)/f_0(x) = c\}$ under either hypothesis is zero. Let

$$(\alpha, \varphi_1), (\alpha, \varphi_2), (\alpha, \varphi_3), \cdots \qquad 0 < \alpha < \varphi_i < 1, \quad i = 1, 2, 3, \cdots, \tag{6.1}$$

be a sequence of pairs of real numbers. Suppose there exists a sequence of sets of the relative optimum zones

$$\begin{gathered}(S_{11}, S_{21}, S_{31}), (S_{12}, S_{22}, S_{32}), (S_{13}, S_{23}, S_{33}), \cdots, \\ (S_{1i}, S_{2i}, S_{3i}) \varepsilon \Omega_{\alpha,\varphi_i,a,r}, i = 1, 2, 3, \cdots,\end{gathered} \tag{6.2}$$

such that the corresponding sequence of R. O. tests will have (6.1) as the sequence of strengths for testing $f_0(x)$ against $f_1(x)$. Let the corresponding sequences of ASN functions be

$$\mu_1(f_i), \mu_2(f_i), \mu_3(f_i), \cdots, \qquad i = 0, 1. \tag{6.3}$$

We shall say that the sequence of R. O. tests is *conditionally consistent*, if for any alternative $f_1(x)$, the sequences of inequalities

$$\mu_1(f_i) < \mu_2(f_i) < \mu_3(f_i) < \cdots, \qquad i = 0, 1, \tag{6.4}$$

imply the sequence of inequalities

$$\varphi_1 < \varphi_2 < \varphi_3 < \cdots. \tag{6.5}$$

This definition is equivalent to

Definition II. The R. O. test is said to be *conditionally consistent*, if, for any fixed level of significance α and any alternative $f_1(x)$, the power function $\varphi(f_1)$ of the R. O. test increases whenever the ASN functions $\mu(f_i)$, for $i = 0$ or 1, increase.

The following two lemmas apply to the R. O. tests.

Lemma 6. *For a fixed level of significance α and fixed alternative $f_1(x)$, the power function $\varphi(f_1)$ is a monotone increasing function of I_0.*

Proof. From (4.2), it is evident that in order to prove Lemma 6, it would be necessary and sufficient to prove that, under the given conditions, if (S_1, S_2, S_3) and (S_1', S_2', S_3') are two different sets of the relative optimum zones such that $I_0' > I_0$, then $A_1'/R_1' < A_1/R_1$. Now, since the level of significance α remains

fixed, then, by (4.2), the equality (5.9a) holds. Therefore, there exists $0 < \rho < 1$ such that

$$A_0' = \rho A_0, \qquad R_0' = \rho R_0. \tag{6.6}$$

By (5.5) and (5.7), these equalities imply that

$$A_1' < \rho A_1, \qquad R_1' > \rho R_1. \tag{6.7}$$

Consequently, we obtain

$$A_1'/R_1' < A_1/R_1, \tag{6.8}$$

which completes the proof of Lemma 6.

LEMMA 7. *For a preassigned level of significance α, the* ASN *functions $\mu(f_i)$ for $i = 0$ or 1 are monotone increasing functions of I_0.*

PROOF. Increasing I_0 decreases S_1 and S_2, and therefore A_0, R_0, A_1 and R_1. Thus, Lemma 7 follows using Lemma 2.

THEOREM 3. *The* R. O. *test is conditionally consistent.*

This theorem follows directly from Lemmas 6 and 7.

Conditional consistency is a rather weak property. It does not assure us that as the average sample number approaches infinity, the power of the R. O. test approaches one. Hence, a stronger property is desirable.

DEFINITION II′. The R. O. test will be said to be *absolutely consistent*, if, for every fixed level of significance α and every given alternative $f_1(x)$, the power function $\varphi(f_1)$ tends to 1 as the ASN function $\mu(f_1)$ tends to ∞.

Although the R. O. test is conditionally consistent, it may not be absolutely consistent. We shall verify this assertion by an example. But first, let us state an obvious but useful lemma.

LEMMA 8. *For a fixed level of significance α, if (S_1, S_2, S_3) is a set of relative optimum zones and if (S_1', S_2', S_3') is any other set of the three zones such that $I_0' = I_0$, then we have*

$$\varphi(f_1;\ a, r, S_1, S_3) \geqq \varphi(f_1;\ a, r, S_1', S_3'). \tag{6.9}$$

This lemma is obviously true by (4.2), (5.5), (5.6), (5.7) and (5.9a).

The following example shows that the R. O. test is conditionally consistent, but not absolutely consistent. Let a class of pdf's be given as follows:

$$f(x) = \theta + 2(1 - \theta)x, \qquad 0 \leqq \theta \leqq 1, \quad 0 < x < 1. \tag{6.10}$$

Let the hypothesis

$$H_0{:}\theta = 1 \tag{6.11}$$

be tested against the alternative hypothesis

$$H_1{:}\theta = \theta_1, \qquad 0 < \theta_1 < 1. \tag{6.12}$$

Clearly, this is equivalent to testing the uniform density $f_0(x) = 1$ against the alternative $f_1(x) = \theta_1 + 2(1 - \theta_1)x$. Since the ratio $f_1(x)/f_0(x) = f_1(x)$ is a mono-

tone increasing function of x, then any set of the relative optimum zones will have the form $S_1 = (0, x)$, $S_2 = (x, x')$, and $S_3 = (x', 1)$. Furthermore, since, for fixed α, both S_1 and S_3 must satisfy the equality

$$R_0 = \lambda A_0, \tag{6.13}$$

where λ is determined so that $\varphi(f_0) = \alpha$, then x and x' must satisfy

$$x' = 1 - \lambda x. \tag{6.14}$$

As a result, we obtain, by (4.2),

$$\beta_1 = 1/(1 + A_1/R_1) = [\lambda(2 - \theta_1) - \lambda^2(1 - \theta_1)x]/ [\theta_1 + \lambda(2 - \theta_1) + (1 - \theta_1)(1 - \lambda^2)x]. \tag{6.15}$$

Taking the limit on β_1 in (6.15), we obtain

$$\lim_{f_0 \to 1} \beta_1 = \lim_{x \to 0} \beta_1 = \lambda(2 - \theta_1)/(\theta_1 + \lambda(2 - \theta_1)) = \beta^* < 1. \tag{6.16}$$

Hence, we have

$$\max_{0 \leqq f_0 < 1} \varphi(f_1) \leqq \varphi^*, \qquad \varphi^* = \int_0^{\beta^*} \frac{(a + r - 1)!}{(a - 1)!(r - 1)!} z^{r-1}(1 - z)^{a-1} \, dz < 1. \tag{6.17}$$

Thus, by Lemma 8, for a given set (α, φ, a, r) with $\varphi > \varphi^*$, the class $\Omega_{\alpha,\varphi,a,r}$ is empty, that is, for the given pair (a, r), there is no set of the three zones giving strength (α, φ) for testing $f_0(x)$ against $f_1(x)$. Therefore, the R. O. test can not be absolutely consistent, though it is always conditionally consistent.

In the following, we give a necessary and sufficient condition for the existence of $\Omega_{\alpha,\varphi,a,r}$ for an arbitrary set (α, φ, a, r) and also a necessary and sufficient condition for the absolute consistency of the R. O. test.

THEOREM 4. *A necessary and sufficient condition for the existence of $\Omega_{\alpha,\varphi,a,r}$ is that there exists a number $k > 0$ such that (A) the probability measure of the set $w_1 = \{x;\ f_1(x)/f_0(x) \leqq k\beta_0/(1 - \beta_0)\}$ is positive under either hypothesis and (B) the probability measure of the set $w_3 = \{x;\ f_1(x)/f_0(x) \geqq k\beta_1/(1 - \beta_1)\}$ is positive under either hypothesis, where $\beta_i = 1/(1 + A_i/R_i)$ for $i = 0$ or 1 are determined from (4.2) so that the R. O. test would have strength (α, φ) for testing $f_0(x)$ against $f_1(x)$.*

PROOF. i) Sufficiency. Since the power functions are continuous under the assumptions, then, from Lemmas 6 and 8, it is clear that in order to prove the existence of $\Omega_{\alpha,\varphi,a,r}$, it would be sufficient to prove the existence of $\Omega_{\alpha,\varphi',a,r}$, where $\varphi' \geqq \varphi$, that is, it is sufficient to show that we can find a set of relative optimum zones (S_1', S_2', S_3') such that the following are satisfied:

$$\text{(a)}\quad A_0'/R_0' = (1 - \beta_0)/\beta_0, \qquad \text{(b)}\quad A_1'/R_1' \leqq (1 - \beta_1)/\beta_1. \tag{6.18}$$

Now, if conditions (A) and (B) hold, we can choose a subset $S_1' \subset w_1$ and a subset

$S_3' \subset w_3$ such that (6.18a) is satisfied. Consequently, we have

$$A_1' \leqq A_0' k\beta_0/(1 - \beta_0), \qquad R_1' \geqq R_0' k\beta_1/(1 - \beta_1). \tag{6.19}$$

Therefore, the inequality (6.18b) holds.

ii) Necessity. Conversely, if the class $\Omega_{\alpha,\varphi,a,r}$ exists, then, by Lemmas 6 and 8, a set of relative optimum zones (S_1'', S_2'', S_3'') exists such that

$$A_0''/R_0'' = (1 - \beta_0)/\beta_0, \qquad A_1''/R_1'' \leqq (1 - \beta_1)/\beta_1 \tag{6.20}$$

hold. Consequently, there exists a number $k > 0$ such that we have

$$A_1''/A_0'' \leqq k\beta_0/(1 - \beta_0), \qquad R_1''/R_0'' \geqq k\beta_1/(1 - \beta_1). \tag{6.21}$$

Therefore, there exist subsets $w_1 \subset S_1''$ and $w_3 \subset S_3''$ such that conditions (A) and (B) are true.

THEOREM 5. *A necessary and sufficient condition for the absolute consistency of the* R. O. *test is that at least one of the following two conditions is true*:

(A′) *for every positive* ϵ, *the probability measure of the set*

$$w_1' = \{x;\ f_1(x)/f_0(x) \leqq \epsilon\}$$

is positive under either hypothesis;

(B′) *for every positive* ϵ', *the probability measure of the set*

$$w_3' = \{x;\ f_1(x)/f_0(x) \geqq \epsilon'\}$$

is positive under either hypothesis.

PROOF. i) Sufficiency. By (4.4), it is seen that the ASN function $\mu(f_1)$ tends to infinity only if at least one of the two quantities A_1 and R_1 tends to zero. Hence, by (4.2), it is obvious that in order to prove the sufficiency it would be sufficient to show that, for every given level of significance α and every alternative $f_1(x)$, the ratio A_1/R_1 tends to zero as R_1 tends to zero. Clearly, for a fixed α, the ratio A_0/R_0 remains fixed. Let

$$A_0/R_0 = d, \tag{6.22}$$

where d is a constant. Then, by (5.7), $R_1 \to 0$ implies $R_0 \to 0$. By (6.22), $R_0 \to 0$ implies $A_0 \to 0$. Finally, by (5.5), $A_0 \to 0$ implies $A_1 \to 0$. Furthermore, if (A′) is true, then

$$\lim_{R_1 \to 0} R_1/R_0 > 1, \qquad \lim_{R_1 \to 0} A_1/A_0 = 0. \tag{6.23}$$

If (B′) is true, then

$$\lim_{R_1 \to 0} R_1/R_0 = \infty, \qquad \lim_{R_1 \to 0} A_1/A_0 < 1. \tag{6.24}$$

Consequently, in either case, we obtain

$$\lim_{R_1 \to 0} A_1/R_1 = \lim_{R_1 \to 0} d(A_1/R_1)(R_0/A_0) = d \lim_{R_1 \to 0} (A_1/A_0)/(R_1/R_0) = 0. \tag{6.25}$$

ii) Necessity. The necessity can be easily proved by contradiction. Assume both (A′) and (B′) are not true. Then, the ratio $f_1(x)/f_0(x)$ must be bounded. Let

$$L = \text{g.l.b. } \{f_1(x)/f_0(x)\}, \qquad U = \text{l.u.b. } \{f_1(x)/f_0(x)\}. \tag{6.26}$$

Choose a set (α, φ, a, r) such that

$$\beta_0/(1 - \beta_0) < L, \qquad \beta_1/(1 - \beta_1) > U, \tag{6.27}$$

where, again, β_0 and β_1 are determined from (4.2) so that the R. O. test should have strength (α, φ). Then, we can not find a $k > 0$ such that (A) and (B) in Theorem 4 hold simultaneously and hence the class $\Omega_{\alpha,\varphi,a,r}$ is empty. This contradicts the assumption that the R. O. test is absolutely consistent. Thus, the necessity of either (A′) or (B′) is established.

From Theorem 5, it is seen that unless (A′) or (B′) is satisfied, the R. O. test can not be absolutely consistent. However, for practical purposes one may modify the procedure and thus obtain a R. O. test with a specified strength (α, φ). The following are two of the possible modifications:

(a) Increasing a and/or r. The power function is in the form of the incomplete beta function. Thus, for an arbitrary pair (α, φ), it may be possible, by increasing a and/or r, to decrease the difference between β_0 and β_1 so that, for some $k > 0$, the conditions (A) and (B) in Theorem 4 are satisfied.

(b) Taking the observations in groups. When observations are taken in groups of size n, one may apply the R. O. test on some appropriate statistic so that the R. O. test will have the specified strength (α, φ). This is because sometimes for some appropriate n, the pdf's of the statistic under the null and the alternative hypotheses may satisfy the condition in Theorem 4. Usually, this is true when n is sufficiently large.

7. Applications. The R. O. test procedure may have a wide variety of applications. In testing a simple hypothesis, the procedure is applicable whenever the pdf under the null hypothesis and the ratio of the pdf's under both the null and the alternative hypotheses are determinable, especially when the condition in Theorem 5 is also satisfied. For example, let $n(x;\ \theta, \sigma^2)$ be the pdf of a normal distribution, that is,

$$n(x; \theta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \theta)^2\right), \qquad -\infty < x < \infty,$$

where σ^2 is known, and let the hypothesis $H_0{:}\theta = \theta_0$ be tested against the alternative hypothesis $H_1{:}\theta > \theta_0$. For $\theta > \theta_0$, the ratio $n(x;\ \theta, \sigma^2)/n(x;\ \theta_0, \sigma^2)$ is a monotone increasing function of x, and both (A′) and (B′) in Theorem 5 are satisfied. Hence we can apply the R. O. test by taking the three intervals $(-\infty, x_1)$, (x_1, x_2), and (x_2, ∞) as S_1, S_2, and S_3, where x_1 and x_2 are determined so that, for fixed "a" and "r", the R. O. test will have a preassigned strength (α, φ) for testing θ_0 against some alternative θ_1 where $\theta_1 > \theta_0$. The determination of x_1 and x_2 can easily be made by trial and error, since x_1 is a monotone decreas-

ing function of φ and x_2 is a monotone increasing function of φ, for a fixed level of significance α. For instance, if $(\alpha, \varphi, a, r) = (.05, .95, 1, 1)$. then, x_1 and x_2 should be so determined that the following two equalities are satisfied:

$$(7.1) \qquad 1/(1 + A_0/R_0) = .05, \qquad 1/(1 + A_1/R_1) = .95,$$

where A_0, A_1, R_0 and R_1 are given by

$$(7.2) \qquad A_i = \int_{-\infty}^{x_1} n(x; \theta_i, \sigma^2)\, dx, \qquad R_i = \int_{x_2}^{\infty} n(x; \theta_i, \sigma^2)\, dx. \qquad i = 0, 1,$$

Thus, if $\theta_1 = \theta_0 + 2\sigma$, then, by the use of the normal probability table, the approximate values of x_1 and x_2 are found to be $x_1 = \theta_0 + .093\sigma$ and $x_2 = \theta_0 + 1.907\sigma$.

If a composite hypothesis is to be tested, sometimes one may also apply the procedure if it is possible to take the observations in groups and a similar region can be found. For instance, the central t-distribution is used in testing the location of the mean of a normal distribution with unknown variance and the χ^2 distribution will be used in testing the variance of a normal distribution with unknown mean.

The following two examples illustrate the application of the test to the nonparametric and multisample problems.

EXAMPLE 1. (Test of the location of the median of a population.) To test whether the median ν of a population is equal to or greater than ν_0, one can take the observations in groups of size n and call an observation 0 if it is less than ν_0 and 1 otherwise. Under the null hypothesis, the sum X of the observations has the binomial density $f(\binom{x}{x}) = \binom{n}{x}(\frac{1}{2})^n$, for $x = 0, 1, 2, \cdots, n$. By grouping the $n + 1$ points $(0, 1, 2, \cdots, n)$ into three different zones, the proposed test is applicable.

EXAMPLE 2. (Comparison of two populations.) Suppose $X_1 < X_2 < \cdots < X_n$ and $Y_1 < Y_2 < \cdots < Y_m$ are the ordered results of two random samples from populations having continuous cumulative distribution functions $F(x)$ and $G(x)$ respectively. Let $s_1, s_2, \cdots, s_n$ be the ranks of the observations of X. Let $W = s_1 + s_2 \cdots + s_n$. Denote by $h(x)$ the pdf of the random variable W. Let the hypothesis H_0:$F(x) = G(x)$ be tested against the alternative hypothesis H_1:$F(x) > G(x)$. Then, since the density $h_0(x)$ of W under H_0 is known, one may apply the test procedure as follows. Choose two positive integers a and r. Decide on two numbers w' and w'' such that

$$(7.3) \qquad \begin{aligned} &\Pr(W < w' \mid H_0) = A_0, \qquad \Pr(w' \leqq W \leqq w'' \mid H_0) = I_0, \\ &\qquad\qquad\qquad\qquad\qquad\qquad \Pr(W > w'' \mid H_0) = R_0, \end{aligned}$$

and such that the pair $(\varphi(f_0), \mu(f_0))$ satisfies certain conditions. Continue to draw samples of sizes (n, m). At each stage, count the number of times that $W < w'$, $w' \leqq W \leqq w''$ and $W > w''$. Denote these numbers by c_1, c_2 and c_3. Then, the proposed test is applicable.

We note that the procedures used in Examples 1 and 2 are not necessarily optimal. They are given here as possible applications of the proposed procedure in general.

8. Efficiency. In this section, we shall investigate the power efficiency of the R. O. test as compared with Wald's sequential probability ratio test.

Let $N(x;\ \theta, \sigma^2)$ be a cumulative normal distribution with an unknown mean θ and known variance σ^2. Let the hypothesis $H_0:\theta = \theta_0$ be tested against an alternative hypothesis $H_1:\theta = \theta_1$. We shall calculate the numerical efficiencies of the R. O. test for the five cases: $\theta_1 = \theta_0 + \lambda\sigma$, $\lambda = 1.0, 1.5, 2.0, 2.5, 3.0$. For each λ, we shall denote by $\psi(\theta)$ and $\eta(\theta)$ the power and the ASN functions of Wald's sequential probability ratio test, and by $\varphi_i(\theta)$ and $\mu_i(\theta)$ the power and ASN functions of the R. O. test for $i = 1, 2$, where by $i = 1$, it is meant $a = r = 1$ and similarly by $i = 2$ is meant $a = r = 2$. Furthermore, let (.05, .95) be the preassigned strength of all the tests, that is, for each $\lambda(\lambda = 1.0, 1.5, 2.0, 2.5, 3.0)$, we have $\psi(\theta_0) = \varphi_i(\theta_0) = .05$ and $\psi(\theta_0 + \lambda\sigma) = \varphi_i(\theta_0 + \lambda\sigma) = .95$ $(i = 1, 2)$. Then, it is obvious that for any real θ, the functions $\psi(\theta)$, $\eta(\theta)$, $\varphi_i(\theta)$ and $\mu_i(\theta)$ $(i = 1, 2)$ depend not only on $\xi = (\theta - \theta_0)/\sigma$, but also on λ (i.e., on H_1). In Tables I and II are given the numerical values of these functions for $\lambda = 1.0, 1.5, 2.0, 2.5, 3.0$ and selected values of ξ. Since the power curves for both the sequential probability ratio and the R. O. tests are close to each other

TABLE I

ξ	Sequential Probability Ratio Test		R. O. Test					
			$a = r = 1$			$a = r = 2$		
	$\psi(\theta)$	$\eta(\theta)$	$\varphi_1(\theta)$	$\mu_1(\theta)$	$\mathcal{E}_1$	$\varphi_2(\theta)$	$\mu_2(\theta)$	$\mathcal{E}_2$
0	.0500	5.2997	.0500	57.4197	.0923	.0500	10.4343	.5079
.50	.5000	8.6695	.5000	117.4618	.0738	.5000	14.9957	.5781
1.00	.9500	5.2997	.9500	57.4197	.0923	.9500	10.4343	.5079
0	.0500	2.3554	.0500	4.2996	.5478	.0500	3.4738	.6780
.50	.2726	3.5711	.2733	7.1090	.5023	.2840	4.4255	.8069
.75	.5000	3.8531	.5000	7.7662	.4961	5000	4.6081	.8362
1.00	.7274	3.5711	.7267	7.1090	.5023	.7160	4.4255	.8069
1.50	.9500	2.3554	.9500	4.2996	.5478	.9500	3.4738	.6780
2.00	.9927	1.5473	.9929	2.5203	.6139	.9951	2.6966	.5738
0	.0500	1.3249	.0500	1.7689	.7490	.0500	2.4051	.5509
.50	.1866	1.8455	.1890	2.3713	.7783	.2018	2.6881	.6865
1.00	.5000	2.1674	.5000	2.7442	.7898	.5000	2.8308	.7656
1.50	.8134	1.8455	.8110	2.3713	.7783	.7982	2.6881	.6865
2.00	.9500	1.3249	.9500	1.7689	.7490	.9500	2.4051	.5509
2.50	.9881	.9581	.9890	1.3671	.7008	.9922	2.1816	.4392
3.00	.9972	.7320	.9979	1.1566	.6329	.9992	2.0649	.3545

TABLE II

ξ	Sequential Probability Ratio Test		R. O. Test $a = r = 1$		$\mathcal{E}_1$
	$\psi(\theta)$	$\eta(\theta)$	$\varphi_1(\theta)$	$\mu_1(\theta)$	
0	.0500	.8479	.0500	1.2241	.6927
.50	.1460	1.1119	.1512	1.4095	.7889
1.00	.3569	1.3484	.3617	1.5770	.8550
1.25	.5000	1.3871	.5000	1.6040	.8648
1.50	.6431	1.3484	.6383	1.5770	.8550
2.00	.8540	1.1119	.8488	1.4095	.7889
2.50	.9500	.8479	.9500	1.2241	.6927
3.00	.9840	.6515	.9862	1.1007	.5919
0	.0500	.5889	.0500	1.0452	.5634
.50	.1232	.7396	.1324	1.0874	.6802
1.00	.2726	.8928	.2860	1.1318	.7888
1.50	.5000	.9633	.5000	1.1519	.8363
2.00	.7274	.8928	.7140	1.1318	.7888
2.50	.8768	.7396	.8676	1.0874	.6802
3.00	.9500	.5889	.9500	1.0452	.5634
3.50	.9807	4718	.9846	1.0185	.4632
4.00	.9927	.3868	.9961	1.0060	.3845

in all the cases considered, then

$$\mathcal{E}_1 = \eta(\theta)/\mu_1(\theta), \qquad \mathcal{E}_2 = \eta(\theta)/\mu_2(\theta), \tag{8.1}$$

as given in the tables can be regarded as the approximate power efficiencies.

From Tables I and II, we observe the following:

(a) In order to obtain high efficiencies, it seems that, when both types of error are fixed and the difference $(\theta_1 - \theta_0)/\sigma$ is small, one should make a and r large.

(b) Some of the figures in the tables are misleading. It is clearly true that no matter which procedure is used, one has to take at least one observation before a decision can be made. Hence, the ASN in either case must be at least one. However, some of the figures for Wald's case are less than one, which can not be regarded as practical. Therefore, in the case $\theta_1 = \theta_0 + 3\sigma$, the efficiencies will be at least .87 uniformly if we assume that ASN is at least one.

(c) If one is interested in improving the efficiency, say, for testing the hypothesis $H_0{:}\theta = \theta_0$ against the alternative hypothesis $H_1'{:}\theta = \theta_0 + \frac{1}{2}\sigma$, then one may take the observations in groups of size 25 and apply the R. O. test on the means $\bar{x}$ (using $a = r = 1$). In other words, one is now testing the same null hypothesis H_0 against an equivalent alternative hypothesis $H_1''{:}\theta = \theta_0 + 2.5\sigma_{\bar{x}}$. Consequently, the efficiencies are raised to at least 69 per cent for all alternatives θ between θ_0 and $\theta_0 + \frac{1}{2}\sigma$.

The author wishes to thank the referee for the helpful comments.

REFERENCES

[1] P. R. HALMOS, *Measure Theory*, D. Van Nostrand Co., Inc., N. Y., 1951.

[2] J. NEYMAN AND E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London Ser. A*, Vol. 231 (1933), pp. 289-337.

[3] A. WALD, "Sequential tests of statistical hypotheses," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 117-186.

[4] A. WALD, *Sequential Analysis*, John Wiley and Sons, Inc., N. Y., 1948.

[5] D. V. WIDDER, *Advanced Calculus*, Prentice-Hall, 1947.

ON A CONTAGIOUS DISTRIBUTION

By R. S. G. Rutherford

University of Sydney, Australia

1. Summary. The purpose of this paper is to discuss the probability distribution that arises when the probability of success at any trial depends linearly upon the number of previous successes. Such a scheme has obvious uses in both biological and economic fields.

It will be shown that by assuming a simple linear relationship between the number of previous successes and the probability of success in the next trial, we can derive a distribution that is reasonably easy to handle, provides as good a fit as more usual distributions, and has parameters which are capable of easy physical interpretation. Moreover, for appropriate values of the parameters the negative binomial and the Gram-Charlier systems can be shown to be close approximations.

2. Introduction. Considerable attention has recently been directed to models where previous experience determines the probabilities in the forthcoming trial. This study is particularly indebted to the work of Woodbury [1]. Much of the recent work has developed the probability scheme originally postulated by Polya [2]. Here it is intended to extend that suggested by Woodbury, and it may be well to contrast the two schemes.

In the Polya scheme, we have an urn containing b black and w white balls. After each random drawing, the drawn ball is returned together with c balls of the same colour. Thus the chance of drawing a ball of given colour depends upon both the number of previous successes and of previous failures.

The Woodbury scheme involves the return of the drawn ball only, if the draw be a failure, and in the event of the draw being a success, the reconstitution of the urn, for example, by the replacement of "failure" balls by "success" balls. In this scheme the order of success is important; in the Polya scheme it is not.

Formally the Woodbury scheme involves that if $P(n, x)$ be the probability of exactly x successes in n trials, and p_x be the probability of success after x previous successes, then

$$P(n + 1, x + 1) = p_x P(n, x) + q_{x+1} P(n, x + 1). \tag{1}$$

Woodbury has solved this problem in the general case.

In this article we postulate further that p_x is a simple linear function of x, viz:

$$p_x = p + cx, \qquad 0 \leq x \leq n. \tag{2}$$

Since we must have $0 < p < 1$, we have the limiting conditions,

$$c > 0, \quad n < q/c; \qquad c < 0, \quad n < p/|c|. \tag{3}$$

Received 6/6/52, revised 5/3/54.

This involves that c must always be of order n^{-1} or smaller. These conditions do not prove very restrictive.

3. The distribution and its properties. Following Woodbury, the solution of (1) and (2) may be shown to be

$$P(n, x) = \frac{1}{x!}\left[\left(\frac{p}{c}\right)\left(\frac{p}{c}+1\right)\left(\frac{p}{c}+2\right)\cdots\left(\frac{p}{c}+x-1\right)\right]\sum_{r=0}^{x}(-1)^r \cdot \binom{x}{r}(q-cr)^n. \tag{4}$$

The summation term is clearly the coefficient of $\theta^n/n!$ in $e^{q\theta}(1 - e^{-c\theta})^x$.

It is desirable to consider the effect of the restrictions of the conditions (2) and (3) upon the value of $P(n, x)$. It will be shown now (a) that $P(n, x)$ is zero for $x > n$, and (b) that $P(n, x)$ is always positive for $0 \leqq x \leqq n$.

(a) Since the term $(1 - e^{-c\theta})^x$ if written as $(c\theta - c^2\theta^2/2! + \cdots)^x$ contains only terms of order θ^x and higher powers of θ, the summation term is zero for all $x > n$.

(b) The condition that $P(n, x)$ as given by relation (4) is always positive within the range $x \leqq n$ requires that with $c > 0$ (i.e., with the product term always positive) the summation term should always be positive, while with $c < 0$ (i.e., with the product term alternately positive and negative) the summation should also alternate in sign, being positive when x is even and negative when x odd. Regarding the summation as the leading xth difference of the series q^n, $(q - c)^n, \cdots, (q - nc)^n$, shows immediately that (3) is a necessary and sufficient guarantee for the summation term to have the correct sign.

The generating function of $P(n, x)$ is the coefficient of $\theta^n/n!$ in $e^{q\theta}[1 - (1 - e^{-c\theta})c]^{-p/c}$ which may be written

$$e^{\theta}[(1 - t)e^{c\theta} + t]^{-p/c}. \tag{5}$$

Though this expression has an infinite number of terms, the terms containing θ^n will occur only in the $n + 1$ terms containing powers of t from 1 to t^n. Thus the generating function is a finite one. That the sum of all the $P(n, x)$ for $x = 0 \cdots n$ is equal to 1 may be confirmed by putting $t = 1$ in (5) and considering the coefficient of $\theta^n/n!$ in e^{θ}. This gives us, for the factorial moment generating function, the coefficient of $\theta^n/n!$ in

$$e^{\theta}[1 - \alpha(e^{c\theta} - 1)]^{-p/c}. \tag{6}$$

Denoting the rth factorial moment by f_r, we have then

$$\begin{aligned} f_1 &= (p/c)[(1 + c)^n - 1], \\ f_2 &= (p/c)(p/c + 1)[(1 + 2c)^n - 2(1 + c)^n + 1], \\ f_3 &= (p/c)(p/c + 1)(p/c + 2)[(1 + 3c)^n - 3(1 + 2c)^n + 3(1 + c)^n - 1]. \end{aligned} \tag{7}$$

4. Empiric fitting. For empiric fitting these three moments (7) should be enough to determine the three parameters n, p, and c. Since the present writer

has been unable to derive a method of fitting on maximum likelihood principles, a somewhat cumbersome method of solution is offered. We may write (7) in the form

$$\frac{p}{c} = \frac{f_1}{(1+c)^n - 1} = \alpha_1 f_1,$$

$$\frac{p}{c} + 1 = \frac{f_2}{f_1}\,\frac{(1+c)^n - 1}{(1+2c)^n - 2(1+c)^n + 1} = \alpha_2 \frac{f_2}{f_1},$$

$$\frac{p}{c} + 2 = \frac{f_3}{f_2}\,\frac{(1+2c)^n - 2(1+c)^n + 1}{(1+3c)^n - 3(1+2c)^n + 3(1+c)^n - 1} = \alpha_3 \frac{f_3}{f_1}.$$

To obtain an estimate of n we can approximate these further as

$$np = [1 - (n-1)c/2]f_1,$$

$$(n-1)(p+c) = [1 - (n-3)c/2]f_2/f_1,$$

$$(n-2)(p+2c) = [1 - (n-5)c/2]f_3/f_1.$$

From these relations p and c may readily be eliminated, giving a cubic for n. In this cubic, $n = 1$ is always a root, and the relation may be reduced to a quadratic, of which the positive root is the only relevant one. Since n must be integral, the nearest integer may be taken as a trial value. Having obtained n, it is easy to evaluate p and c. The terms of the distribution are very sensitive to small changes in p and c, which should be evaluated carefully.

It is intended to develop tables of the values of the expressions α_1, α_2, and α_3 which will make the fitting less arduous, and more reliable for ranges in which the above approximations are not valid.

5. Comparisons. The results of fitting this distribution to two classical sets of data are given in Tables I and II. In both cases, the fit of the present distribution is at least as good as in the standard fittings. The improvement is not remarkable, but the parameters of the distribution have a clear physical meaning which can never be claimed for the parameters of the negative binomial or the Neyman contagious set [5]. That is the major claim made for this work.

It is intended now to investigate why other distributions appear to be close approximations in certain circumstances. It is important, however, to make plain the purpose of the following sections. There is no intention to discuss the

TABLE I

Accidents to women working on H.E. shells, data of Greenwood and Yule [3]

$n = 6 \qquad p = .059886 \qquad c = 0.103036$

Number of accidents	0	1	2	3	4	5	Tot.
Observed frequency	447	132	42	21	3	2	647
Negative binomial	442	140	45	14	5	2	648
Neyman contagious distribution	448	128	49	16	5	1	647
Present distribution	447	130	47	17	5	1	647

TABLE II

Yeast cells in 400 squares of a haemcytometer, data of "Student" [4]

$n = 13 \qquad p = .046747 \qquad c = 0.019088$

Number of yeast cells	0	1	2	3	4	5	Tot.
Observed frequency	213	128	37	18	3	1	400
Negative binomial	214	123	45	13	4	1	400
Present distribution	215	122	45	14	3	1	400
Gram-Charlier Type B	216	119	46	15	3	1	400

minutae of the conditions under which the approximations will be valid. Such conditions may be found by anyone sufficiently interested.

The purpose in this context is merely to explain why certain distributions have provided reasonably good fits to empiric data which may have, in fact, been generated by a system of the type of (1). A second, and perhaps subsidiary, point is that the fitting of the distribution is difficult, particularly as no maximum likelihood method seems available. This may be overcome in certain ranges by fitting these other distributions, where the parameters are easier to determine, if these parameters can be interpreted in terms of those of the present distribution.

To illustrate and confirm the following sections, a number of actual distributions have been evaluated, together with the approximations under discussion. These are given in Section 8.

6. Binomial approximations. In the negative binomial generated by

$$[(1 + P) - Pt]^{-k}, \tag{8}$$

we have, by standard methods

$$f_1 = kP, \qquad f_2 = k(k + 1)P^2, \qquad f_3 = k(k + 1)(k + 2)P^3. \tag{9}$$

By the method of moments we can then determine the parameters as

$$k = f_1^2/(f_2 - f_1^2), \qquad P = (f_2 - f_1^2)/f_1 . \tag{10}$$

Comparing (9) with (7) shows a considerable similarity of form, if we assume that c and hence p/c are positive. If c and n be small enough for us to equate $(1 + 2c)^n$ and $(1 + c)^{2n}$, we will have at once

$$k = p/c, \qquad P = (1 + c)^n - 1. \tag{11}$$

The necessary conditions for this to be valid are somewhat complicated but involve that terms in n^2c^2 may be neglected and/or that $p/c < n - 1$. In practice the first of these is rarely likely to be obtained. However, it can still be demonstrated, by some rather cumbersome analysis not shown here, that so long as $p/c < n - 1$ a negative binomial can be fitted, though the parameters no longer bear easy interpretation in terms of those of the original distribution.

A positive binomial generated by, say, $(Q' + P't)^{k'}$ might provide a good fit if c is either negative or positive with $p/c \geq n - 1$, that is with $f_1^2 > f_2$. The

first possibility is restricted by the fact that it also would appear to require strictly $p/|c| < n - 1$, contrary to (3). The approximation, however, is reasonably good for values of $p/|c|$ of about the same order as n. The second case also seems to be relevant only when p/c exceeds $n - 1$ by only a small amount. Cases where p/c greatly exceeds $n - 1$ may be handled more satisfactorily otherwise, as in the following section.

Both cases, however, suffer from the difficulty that the value of k' is not, in general, integral. This will not be an insuperable difficulty if P' be small and k' large, for then we may be in the territory where a Poisson distribution may approximate to the positive binomial and hence to the original distribution. However, again it seems that large values of p/c or $p/|c|$ (whether exceeding n or not) may be dealt with best by the Gram-Charlier approximations.

7. Gram-Charlier approximations. *Stage I, binomial type.* Let us now consider cases where the ratio p/c is large. Returning to (4), we have already shown that the summation term is the coefficient of $\theta^n/n!$ in $e^{q\theta}(1 - e^{-c\theta})^x$. If c be sufficiently small, then we have

$$(1 - e^{-c\theta})^x/c^x = \theta^x(1 - c\theta x/2 + c^2\theta^2 x(3x + 1)/24 - \cdots). \tag{13}$$

Hence the summation term is

$$\frac{n!}{n - x!}c^x q^{n-x}\left(1 - \frac{cx(n - x)}{2q} + \frac{c^2x(3x + 1)(n - x)(n - x + 1)}{24q^2} - \cdots\right) \tag{14}$$

Alternatively, with c small, the summation term is

$$\begin{aligned}\Delta^x(q - cx)^n &= \frac{d^x}{d\theta^x}\bigg|_{\theta=x/2}(q - c\theta)^n = \frac{n!}{n - x!}\left(q - \frac{cx}{2}\right)^{n-x}c^x \\ &= \frac{n!}{n - x!}c^x q^{n-x}\left(1 - \frac{cx(n - x)}{2q} + \frac{c^2x^2(n - x)(n - x + 1)}{8q^2}\right).\end{aligned} \tag{14a}$$

These approximations are, of course, true for all values of p/c.

If we leave the product part of $P(n, x)$ in its original form, obviously we can obtain as an approximation to the whole expression

$$P(n, x) = \binom{n}{x}p(p + c)(p + 2c)\cdots[p + (x - 1)c](q - cx/2)^{n-x}. \tag{15}$$

In this form there is more hope of a maximum likelihood fit.

If, however, p/c be sufficiently large, we may write the product term of (4) as

$$\begin{aligned}&\frac{p(p + c)\cdots[p + (x - 1)c]}{x!c^x} \\ &= \frac{p^x}{x!c^x}\left(1 + \frac{c}{p}\sum_{r=0}^{x-1} r + \frac{c^2}{p^2}\sum_{r=0}^{x-1}\sum_{s=0}^{x-1} rs\right) \\ &= \frac{p^x}{x!c^x}\left(1 + \frac{c}{p}\frac{x(x - 1)}{2} + \frac{c^2}{p^2}\cdot\frac{x(x - 1)(x - 2)(3x - 1)}{24}\right),\end{aligned} \tag{16}$$

and hence derive

$$P(n, x) = \binom{n}{x} q^{n-x} p^x \Big\{ 1 + x[x - (np + q)] \frac{c}{2pq}$$

$$(17) \qquad + x\,[p^2(3x + 1)(n - x)(n - x + 1) - 6pqx(x - 1)(n - x)$$

$$+ q^2(x - 1)(x - 2)(3x - 1)] \frac{c^2}{24p^2q^2} - \cdots \Big\}.$$

The term in c^2 will be at most of order $c^2n^4/8$, and subsequent terms of smaller order. If, therefore, such terms may be neglected, we may write

$$(17a) \qquad P(n, x) = \binom{n}{x} q^{n-x} p^x \left[1 + x[x - (np + 8)] \frac{c}{2pq} \right].$$

It is instructive to compare this distribution with the binomial distribution having constant probability p. With c positive, that is, probability increasing, $P(n, x)$ exceeds the corresponding binomial term for $x > np + q$, and $P(n, x)$ is less than the corresponding binomial term for $0 < x < np + q$; with c negative, the conditions are reversed. (It can also be established that the conditions on c that make the approximations valid also ensure that $P(n, x)$ is always positive.) If, moreover, n and p are of the order to make the binomial symmetrical, the skewness of the distribution is an immediate guide to the sign and magnitude of c.

Stage II-A, Limiting form for large n and large p. As indicated in Section 5, the conditions necessary to ensure that the approximations will be valid for all x have not been elaborated. It is immediately obvious that much less stringent conditions will apply for early terms of the distribution than those required for the whole distribution. For central values of p, the latter terms of the binomial part of the expression for the distribution will in any case be small, and the absolute if not the proportionate error small.

These considerations become important when we examine the limiting form of (17) when n becomes large. By the change of variable $X = (x - np)/\sqrt{npq}$ used to transform the binomial into the normal distribution, we find as the continuous distribution parallel to the normal

$$(18) \qquad dP = \phi(X)[1 - \tfrac{1}{2}nc + \{n(n - 1)pc \,/\, 2\sqrt{npq}\}X + \tfrac{1}{2}ncX^2]\, dX,$$

where $\phi(X) = (2\pi)^{-1/2} \exp\{-\tfrac{1}{2}X^2\}$. If c/p be not quite small enough to make the Stage I approximations valid, there will be discrepancies at the right tail. The fit will be poor at both tails in any case, in the same way as the normal is a poor approximation to the binomial at the tails. But, by and large we may expect to get a good fit with a curve of the form

$$(19) \qquad dP = \phi(X)[(1 - a_2) + a_1X + a_2X^2]\, dX.$$

By transferring the origin to the mean $x = a_1$ and standardising the distribution, we may obtain readily

$$(20) \qquad dP = \phi(X)[1 + \mu_3 H_{(3)}/3! + (\mu_4 - 3)H_{(4)}/4! + \cdots]\, dX,$$

which is the standard Gram-Charlier Type A distribution [6]. As shown earlier, the skewness of the system indicates the type of scheme operating, that is, the sign and relative magnitude of c. Consideration of the order of the terms involved indicates that we need consider only the terms listed.

Stage II-B, Limiting form for large n and small p. The limiting process used in Stage II-A is of course valid only if p is not small. It is interesting to investigate whether with p small (but p/c still large), we obtain the Gram-Charlier Type B distribution. For all p we may write (17) in the form

$$(21)\quad P(n, x) = \binom{n}{x} q^{n-x}p^x - \lambda\left[\binom{n-1}{x-1} q^{n-x}p^{x-1} - \binom{n-2}{x-2}q^{n-x}p^{x-2}\right],$$

where $\lambda = \frac{1}{2}n(n-1)(p/q)c$. If now p be small, and of order n^{-1}, such that

$$np = m_1, \qquad (n-1)p = m_2, \qquad (n-2)p = m_3,$$

when n becomes large we obtain

$$(22)\qquad P(n, x) = e^{-m_1}m_1^x/x! - \lambda e^{-m_2}m_2^{x-1}/(x-1)! + \lambda e^{-m_3}m_3^{x-2}/(x-2)!.$$

It is now reasonable to equate the m's and write

$$P(x) = e^{-m}[m^x/x! - \lambda m^{x-1}/(x-1)! + \lambda m^{x-2}/(x-2)!],$$

which again may be written

$$(23)\qquad P(x) = (e^{-m}m^x/x!)[1 - \lambda x/m + \lambda x(x-1)/m^2].$$

This is the required Gram-Charlier Type B [6]. It is most easily fitted by means of the relations

$$(24)\qquad \mu_1' = m + \lambda, \qquad \mu_2' = m + 3\lambda - \lambda^2,$$

which can be solved readily for m and λ. As an illustration, "Student's" data have been fitted by this distribution also (Table II). We have $m = 0.61567$ and $\lambda = 0.06683$. The fit, though reasonably good, is poorer than those previously considered; this may reasonably be attributed to the relatively low values of $n = 13$ and of $p/c = 2.45$.

8. Calculations. The validity of the approximations suggested above is demonstrated by 16 examples in Table III. We have here a number of distributions calculated with a selection of values for n, p, and c, and the values given by the relevant approximating distributions. All values are quoted to four decimals, though they have been calculated to five or more. The approximating distributions used are identified by roman numerals.

Type I is the negative binomial fitted from the moments of the distribution. In each case the values of the parameters used are given for comparison with those given by (11), which also are given.

Type II is the approximation of the form suggested in (17a). In this case there is no attempt to find n, p, and c from the data to give the closest fitting curve of this type; the values used are those of the original distribution. Table III shows

TABLE III

Comparison of exact probabilities with those of various approximations (denoted by roman numerals as explained in Section 8) for different values of n, p, and c. Final entry in each row is for vales of $x \geqq$ that at head of column, except that entries in $x = 10$ column are for that value only. Asterisk (*) indicates value is $< .00005$.

n	p	c	$p/\|c\|$	x:	0	1	2	3	4	5	6	7	8	9	10	Parameters
GROUP A																
100	.0005	.005	0.10	Exact:	.9512	.0375	.0081	.0022	.0007	.0003						
				I:	.9513	.0375	.0081	.0022	.0007	.0002						$P = 0.642400$ $k = 0.100655$
100	.0025	.005	0.50	Exact:	.7786	.1538	.0453	.0147	.0050	.0017	.0006	.0002	.0001			
				I:	.7783	.1543	.0451	.0146	.0050	.0017	.0006	.0002	.0001			$P = 0.631132$ $k = 0.512310$
100	.0050	.005	1.00	Exact:	.6058	.2397	.0943	.0369	.0143	.0055	.0020	.0008	.0007			
				I:	.6052	.2407	.0942	.0366	.0142	.0055	.0021	.0008	.0007			$P = 0.625914$ $k = 1.033160$
100	.0500	.005	10.00	Exact:	.0059	.0243	.0544	.0882	.1156	.1295	.1289	.1167	.3426			
				I:	.0056	.0243	.0568	.0932	.1229	.1381	.1375	.1247	.2969			$P = 0.532497$ $k = 12.1441$
				II:	.0059	.0230	.0485	.0838	.1073	.1383	.1561	.1510	.2861			$np + q = 5.95$ $c/2pq = 19$
GROUP B																
20	.0010	.010	0.10	Exact:	.9802	.0179	.0017	.0002	*							
				I:	.9801	.0180	.0017	.0002	*							$P = 0.200494$ $k = 0.100824$
20	.0050	.010	0.50	Exact:	.9046	.0827	.0109	.0015	.0002	.0001						
				I:	.9043	.0843	.0106	.0015	.0002	*						$P = 0.193605$ $k = 0.568657$
20	.0100	.010	1.00	Exact:	.8179	.1503	.0265	.0045	.0007	.0001						
				I:	.8170	.1519	.0259	.0043	.0007	.0002						$P = 0.184377$ $k = 1.119426$
20	.1000	.010	10.00	Exact:	.1216	.2435	.2574	.1894	.1080	.0505	.0296					
				I:	.1136	.2442	.2653	.1902	.1055	.0473	.0339					$P = 0.023194$ $k = 94.88$
				II:	.1216	.2405	.2567	.1933	.1104	.0492	.0283					$np + q = 2.9$ $c/2pq = 18$
				V:	.1353	.2420	.2420	.1804	.1093	.0551	.0359					$m = 2$ $\lambda = 0.211111$
GROUP C																
10	.1000	.010	10.00	Exact:	.3487	.3686	.1950	.0674	.0169	.0034						
				II:	.3487	.3680	.1953	.0677	.0168	.0033						$np + q = 1.9$ $c/2pq = 18$
				IV:	.3513	.3675	.1922	.0686	.0195	.0029						$m = 1.0462$
				V:	.3679	.3485	.1840	.0705	.0214	.0077						$m = 1.00$ $\lambda = 0.05$

10	.2000	.010	20.00	Exact:	.1074	.2538	.2870	.2039	.1002	.0477						$np + q = 2.8$ $c/2pq = 32$
				II:	.1074	.2533	.2859	.2051	.1012	.0471						
10	.7000	.010	70.00	Exact:	*	.0001	.0011	.0066	.0266	.0770	.1625	.2467	.2583	.1687	.0523	$np + q = 7.3$ $c/2pq = 42$
				II:	*	.0001	.0011	.0062	.0252	.0747	.1629	.2535	.2646	.1652	.0464	$X = .69x - 4.83$ $a_1 = .217$ $a_2 = .05$
				III:	*	.0001	.0007	.0050	.0255	.0821	.1765	.2569	.2401	.1436	.0695	
GROUP D																
10	.5000	.010	50.00	Exact:	.0010	.0089	.0382	.1007	.1809	.2317	.2142	.1411	.0634	.0176	.0023	$np + q = 5.5$ $c/2pq = 50$
				II:	.0010	.0089	.0378	.0996	.1805	.2338	.2174	.1418	.0615	.0159	.0018	$X = .6325x - 3.1625$ $a_1 = .1423$ $a_2 = .050$
				III:	.0022	.0099	.0367	.0982	.1765	.2361	.2113	.1382	.0692	.0175	.0042	
10	.5000	.020	25.00	Exact:	.0010	.0082	.0332	.0862	.1579	.2138	.2167	.1625	.0863	.0294	.0049	$np + q = 5.5$ $c/2pq = 25$
				II:	.0010	.0080	.0316	.0820	.1559	.2215	.2297	.1664	.0791	.0221	.0027	$X = .6325x - 3.1625$ $a_1 = .2846$ $a_2 = .100$
				III:	.0022	.0091	.0312	.0803	.1558	.2272	.2268	.1618	.0745	.0250	.0061	
10	.5000	.040	12.50	Exact:	.0010	.0069	.0252	.0626	.1170	.1723	.2030	.1899	.1359	.0679	.0183	$np + q = 5.5$ $c/2pq = 12.5$
				II:	.0010	.0063	.0193	.0469	.1066	.1969	.2543	.2156	.1143	.0344	.0045	$X = .6325x - 3.1625$ $a_1 = .5692$ $a_2 = .200$
				III:	.0022	.0069	.0189	.0465	.1082	.2034	.2502	.2095	.1055	.0487	.0100	
GROUP E																
10	.1000	−.010	10.00	Exact:	.3487	.4074	.1906	.0464	.0065	.0004						$np + q = 1.9$ $c/2pq = -18$
				II:	.3487	.4068	.1910	.0467	.0059	.0009						$m = .9562$
				IV:	.3844	.3675	.1757	.0560	.0131	.0033						
10	.5000	−.010	50.00	Exact:	.0010	.0107	.0505	.1359	.2290	.2559	.1898	.0926	.0284	.0049	.0004	$np + q = 5.5$ $c/2pq = -50$
				II:	.0010	.0106	.0501	.1348	.2297	.2584	.1928	.0926	.0264	.0036	.0001	$X = .6325x - 3.1625$ $a_1 = .1423$ $a_2 = .050$
				III:	.0022	.0121	.0487	.1322	.2227	.2711	.1879	.0922	.0262	.0045	.0002	

the constant $np + q$ and $c/2pq$, used in (17a) with the corresponding values of n and p.

Type III presents the areas cut off in the range $x \pm \frac{1}{2}$ of the continuous distribution of the form of (19). Table III shows the transformation from x to X, and the values of a_1 and a_2 as calculated from the initial values of n, p, and c, with no effort at improvement.

Type IV is the positive binomial, used where the moments make it impossible to fit a negative binomial. Since the exponent is not an integer, it is fitted as a Poisson, of which the parameter m is given.

Type V is a Gram-Charlier Type B of the form of (23), fitted when appropriate. The parameters m and λ, obtained from (24), are given.

The 16 examples fall into five groups, which examine different aspects of the approximations.

Group A has $n = 100$ and $c = .0050$ throughout. The value of p varies from .0005 to .0500 and the ratio p/c from 0.1 to 10.0. With n as high as 100, only the early terms can be evaluated readily. This limits the possible range of p, and of the ratio p/c. The negative binomial provides a very good fit for low values of p/c. It is still good for the earlier terms of the fourth example, and on the whole better than the Type II approximation.

Group B has a smaller n of 20 and a larger c of 0.01, but values of p such that the same four values of p/c are obtained. Again, for low values of p/c the negative binomial gives a very good fit. For larger values of this ratio, however, Types II and V are also relatively good fits.

Group C comprises three examples where p/c is larger than n, which is 10 in all three. The negative binomial can no longer be fitted. The Type II approximations are reasonably good in all three cases. The Type IV and V approximations in the first example suffer from the small values of n, while the poor Type III approximation in the third example reflects the poorness of the normal as an approximation to the binomial with $n = 10$ and p as large as 0.70.

Group D presents three examples designed to examine the validity of the Gram-Charlier Type A approximation, Type III. Since n is small, a limitation produced by the practical difficulties of computing the "exact" series, central values of p have been taken because the binomial-normal approximation is closest at these values. In the first two examples the Type II fit it sufficiently good to make the Type III fit reasonable. In the third example, with a larger $c = 0.04$, the Type II fit is relatively poor and the Type III fit is worse, reflecting the fact that terms in c^2 may no longer be neglected in (17).

Group E contains two examples in which c is negative, -0.01, with n still 10. Types II and IV are used in the first example, with $p = 0.10$, and Types II and III in the second, with $p = 0.50$.

9. Significance of the results. In all fields of scientific investigation the end goal is always explanation rather than mere description. The negative binomial and the Gram-Charlier set have been found to be good descriptive fits for a large

number of empiric distributions. Probability systems to "explain" them have also been available.

The assumption that we are sampling from a population where the probability varies between individual members and is distributed in the form of a gamma variate will produce as the expected distribution the negative binomial. Equally, the Gram-Charlier system may be derived as the resultant of a small number of linearly additive independent causes of about the same order of importance.

In both cases, distributions arise where these explanations are unconvincing. It has been shown above that a much simpler hypothesis will produce distributions that are at least as good a fit, and which in some cases, though perhaps not in all, provides a more convincing "explanation." As with the normal distribution, we can choose which of two alternative "explanations" is most suitable in any particular case.

The fact that the same probability scheme "explains" both types of distribution considerably systematises the field. Moreover, it appears that there are large areas of possible values for the parameters n, p, and c, where the approximations will not be valid. It is hoped that many empiric distributions which previously have appeared to obey no simple law now may become more tractable.

It is possible, with reasonable ease, to establish that both the negative binomial and the Gram-Charlier Type B distributions may be considered as special cases of the Neyman contagious distribution, and hence that our present distribution will often be closely represented by it. It is more difficult to establish a direct connection, but other writers may succeed.

Though Neyman claims for his series that "All the constants introduced have meanings which are easy to interpret," this does not appear to have been general experience. The distribution of the present study may be of equally general application and provide opportunities for much simpler interpretation.

REFERENCES

[1] M. Woodbury, "On a probability distribution," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 311–313.

[2] G. Polya, "Sur quelques points de la théorie des probabilités," *Ann. Inst. H. Poincaré*, Vol. 1 (1931), pp. 117–162.

[3] M. Greenwood and G. U. Yule, "An enquiry into the nature of frequency distributions of multiple happenings, etc.," *J. Roy. Stat. Soc.*, Vol. 83 (1920), pp. 255–304.

[4] "Student," "On the error of counting with a Haemacytometer," *Biometrika*, Vol. 5 (1907), pp. 351–364.

[5] J. Neyman, "On a new class of contagious distributions applicable in entomology and bacteriology," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 35–57.

[6] M. Kendall, *Advanced Theory of Statistics*, Vol. 1., Griffen and Co., 1945, pp. 145–155.

IDENTIFICATION AND ESTIMATION OF LINEAR MANIFOLDS IN n-DIMENSIONS[1]

By T. A. Jeeves

University of California, Berkeley

1. Summary. This paper investigates the problem of identifiability and estimability of linear structures in n dimensions. The concept of identifiability is examined to elucidate the senses in which it may be interpreted in the present problem. Particular attention is given to the question of treating linear subspaces rather than specific coordinate systems. Necessary and sufficient conditions for identifiability are obtained under the assumption that the "errors" follow a multinormal distribution.

2. Introduction. In many fields of statistical application it is not possible to observe directly the variables of interest but only to observe related random variables. Let $X = (X_1, X_2, \cdots, X_n)$ be a random (row) vector which is "unobservable" and $Y = (Y_1, Y_2, \cdots, Y_n)$ be a random (row) vector which is "observable." Assume that $Y = X\mathbf{B} + U$, where $\mathbf{B}$ is a parameter having $n \times n$ matrices of sure numbers for values and $U = (U_1, U_2, \cdots, U_n)$ is a random (row) vector which is stochastically independent of X.

In this paper particular attention will be given to the case in which U has a multinormal distribution, and it is desired to determine the row space S of the value of $\mathbf{B}$. Two problems are considered: (a) identifiability, whether S is determined if the distribution of Y is known [1], [2], and (b) estimability, whether S can be estimated consistently [1] from an infinite sequence of observations on Y.

Similar problems were considered in 1901 by Pearson [3]. As early as 1916, Thomson [4] showed that estimates based on moments no higher than the second would not be consistent. In 1936, Neyman [5] indicated a set of conditions in which, because of nonidentifiability, no consistent estimates existed. A summary of the state of the problem in 1940 was given by Wald [6], who brought an entirely new approach. An answer for the case of two dimensions was supplied by Reiersøl [7] in 1948.

3. Identifiability. The problem of identification in n dimensions introduces features not present in the two dimensional problem. In particular, just what is to be identified and hence estimated must be clarified. In n dimensions a greater variety of possible interpretations is available. To elucidate the sense in which the problem is treated here, and to bring out the relationships to other work, it seems necessary and profitable to begin with some general remarks on identification, culminating in the definition of identifiability (Definition 3) utilized in the remainder of the paper.

Received 12/15/52, revised 2/12/54.

[1] This paper was prepared with the partial support of the Office of Naval Research.

Following Neyman [1], the concept of identifiability is defined in the following way. Let L be a relation $L(\vartheta, F)$ between the elements ϑ of a space Θ and the elements F of a set Ω of distribution functions. Let $\omega(\vartheta) = \{F \mid L(\vartheta, F)\}$. Let θ be a parameter with range $\Theta_s \subset \Theta$.

DEFINITION 1. θ is *identifiable* (L) if the sets $\omega(\vartheta)$ are disjoint for every $\vartheta \in \Theta_s$.

This definition generalizes that of Neyman, in that Θ and Θ_s are not necessarily identical. The definition emphasizes the relation L between elements of Θ and elements of Ω. Essentially the definition states that identifiability obtains if no two distinct elements of Θ_s are related to the same distribution function. However, for the succeeding discussion, it is important to notice that this definition implies the existence of a relation among the elements of the larger space Θ, and that this relation characterizes the nature of the identification. The following theorem, which follows easily from the definition, brings out this point. The following notation is introduced: for any $F \in \Omega$,

$$\gamma(F) = \{\vartheta \in \Theta \mid F \in \omega(\vartheta)\}, \qquad \Gamma(\vartheta) = \bigcup_{F \in \omega(\vartheta)} \gamma(F), \qquad \Omega_s = \bigcup_{\vartheta \in \Theta_s} \omega(\vartheta).$$

THEOREM 1. *θ is identifiable (L) if and only if there is a relation R between the elements of Θ such that for every $\vartheta \in \Theta_s$ and every $\vartheta^* \in \Gamma(\vartheta)$,*

(i) *$R(\vartheta, \vartheta^*)$ holds, and*

(ii) *$R(\vartheta^*, \vartheta)$ holds only if $\vartheta^* = \vartheta$.*

The relation R is uniquely defined by the relation L for any $\vartheta \in \Theta_s$ and $\vartheta^* \in \Gamma(\vartheta)$; conversely, specifying R implies restrictions on L.

From this it is seen that if θ is identifiable (L) then there is a one-to-one correspondence for $\vartheta \in \Theta_s$ between ϑ and $\omega(\vartheta)$, and also between ϑ and $\Gamma(\vartheta)$. Further, every $F \in \Omega_s$ determines a unique value of ϑ and there exists a function I with the domain Ω_s and range Θ_s such that if $F \in \Omega_s$, then $F \in \omega(I(F))$. If $\Theta = \Theta_s$, then the relation R is equality. In the following, particular attention will be given to the case in which the ϑ are linear spaces and R is the relation of inclusion.

To apply the definition to the problem considered, it is necessary to exhibit the relation L. Let M be the set of $n \times n$ matrices and Θ a family of subsets of M. Let Ω be the set of n-dimensional distribution functions and $\mathfrak{X}$, $\mathfrak{U}$, and $\mathfrak{Y}$ be nonempty subsets of Ω associated with the random variables X, U, and Y, respectively. Let F_X, F_U, and F_Y be the distribution functions of the random variables X, U, and Y, respectively.

DEFINITION 2. For any sets $\mathfrak{X}$, $\mathfrak{U}$, and Θ, the relation $L(\vartheta, F_Y)$ holds if $B \in \vartheta$ and if $Y = XB + U$ for some X and U such that $F_X \in \mathfrak{X}$ and $F_U \in \mathfrak{U}$.

It has been shown [8] that conditions must be imposed on both $\mathfrak{X}$ and $\mathfrak{U}$ if θ is to be identifiable (L).

Further analysis of the problem requires consideration of the effect on the matrix B of a nonsingular transformation P. Identification problems may be proposed in which the space Θ_s is so specialized that PB no longer belongs to an element of Θ_s, or may not for certain P. Such problems will not be considered in this paper. It will be assumed that Θ_s has the following property: for any non-

singular $n \times n$ matrix P, if $B \varepsilon \vartheta \varepsilon \Theta_s$, then $PB \varepsilon \vartheta^*$ for some $\vartheta^* \varepsilon \Theta_s$. Considering the definition of L above it follows that:

THEOREM ?. *In any sets* Θ, Θ_s, $\mathfrak{X}$, *and* $\mathfrak{U}$ *such that* $\Theta_s \varepsilon \Theta$ *and* Θ_s *has the above property, if* θ *is identifiable* (L) *then for each nonsingular matrix* P *either*

(a) $F_X \varepsilon \mathfrak{X}$ *implies* F_{XP} *not a member of* $\mathfrak{X}$ *for every* X, *or*

(b) $B \varepsilon \vartheta$ *implies* $P^{-1}B \varepsilon \vartheta$ *for every* $\vartheta \varepsilon \Theta_s$.

The content of this theorem indicates two broad categories of problems, those in which condition (a) is satisfied by all nonsingular matrices and those in which condition (b) is satisfied by all nonsingular matrices. Mixed problems in which some matrices P satisfy (a) and some (b) might also be considered. The assumption that condition (a) is satisfied for every P leads to the consideration developed by Koopmans [2], [9].

This paper explores the implications of assuming that condition (b) is satisfied for all P, that is, that the matrices belonging to ϑ are all row equivalent or have the same row space. It will thus be convenient to think of ϑ as a row space. With this interpretation the problem being considered below is that of identifying the row space of the matrix B. The row space is a natural parameter in the problem of general linear structures. As such problems frequently arise, the components of X are presumed to lie in a linear subspace of Euclidean n-space; the determination of this linear subspace is desired. The specification of a particular set of coordinates on this subspace (that is, the determination of B) is frequently not required.

Throughout the remainder of this paper it will be assumed that the elements of Θ_s are the sets of row-equivalent matrices corresponding to the various row spaces of dimension s and that $\Theta = \bigcup_1^n \Theta_s$. It will also be assumed that $\mathfrak{U}$ is the set of multinormal distributions. Since $\mathfrak{X}$ is not specified, the relation L is not completely determined. Instead of specifying the set $\mathfrak{X}$, it will suffice to select the relation R (see Theorem 1) and investigate what conditions on $\mathfrak{X}$ are necessary and sufficient for identifiability. Two natural relations among linear spaces are the relation of equality and the relation of inclusion. The treatment here will be confined to the relation of inclusion. Similar results for the relation of equality have been obtained [8].

In view of the preceding considerations, the definition of identifiability may be particularized for the relation of inclusion as follows:

DEFINITION 3. θ is *identifiable* (L^*) if $\mathcal{S}(\vartheta) \subset \mathcal{S}(\vartheta^*)$ for every $\vartheta \varepsilon \Theta_s$ and $\vartheta^* \varepsilon \Gamma(\vartheta)$.

Here $\mathcal{S}(\vartheta)$ denotes the row space of ϑ, that is, the vector space spanned by the row vectors of any element of ϑ, while L^* denotes a relation L which gives rise to the relation R of inclusion (see Theorem 1). Here $R(\vartheta, \vartheta^*)$ means $\mathcal{S}(\vartheta) \subset \mathcal{S}(\vartheta^*)$.

4. Necessary and sufficient conditions. As in the case of two dimensions [7], identifiability is related to a lack of normality in the random variable X. This concept of the amount of nonnormality of a random variable is defined below.

DEFINITION 4. The *dimension* of a random variable U is the smallest dimension of all linear subspaces which contain U with probability one.

DEFINITION 5. Let $nn(Y)$ be the least value of d such that $Y = U + V$ with U and V independent, V having a multinormal distribution and U having dimension d. This value $nn(Y)$ will be called the *nonnormality* of Y.

DEFINITION 6. The nonnormality of $\mathcal{Y}$ is s (i.e., $nn(\mathcal{Y}) = s$) if $nn(Y) = s$ for every Y such that $F_Y \varepsilon \mathcal{Y}$.

In terms of the definition of nonnormality, the main result on identifiability of linear manifolds in n-dimensions can be stated as follows.

THEOREM 3. *θ is identifiable (L^*) if and only if $nn(\mathcal{Y}) = s$.*

The proof of the theorem depends on the following lemmas, the proofs of which are straightforward [8] and will not be given.

LEMMA 1. *If M is an $n \times n$ matrix with rank s and if $(n - s)$ columns of M are identically zero, then every row either belongs to some $s \times s$ submatrix with rank s, or else is identically zero.*

LEMMA 2. *If A is a symmetric matrix, E is a diagonal matrix with ones and zeros on the main diagonal, and EAE is positive semidefinite, then there exist matrices C, G, and H such that*

$$CAC' = DGD + EHE,$$

where $D + E = I$ (the identity) and H is a diagonal matrix with ones and zeros on the main diagonal, C is nonsingular, and $CD = D$.

The following choice of notation has been made. The symbol $f(t)$ will be used to denote some polynomial of the second degree in t, but not necessarily the same polynomial at each usage. Distinct polynomials will not generally be distinguished. The characteristic function of a random variable X will be denoted by $\varphi_X(t) = \int e^{itx'}\, dF_X(x)$, where t and x are row vectors. Further, $\psi_X(t) = -\log \varphi_X(t)$.

LEMMA 3. *If $\psi_Y(t) = \psi_X(tB') + \psi_U(t)$, and if U has a multinormal distribution, then for any matrix C which is idempotent and row-equivalent to B,*

$$\psi_Y(t) = \psi_Y(tC') + f(t).$$

In particular C may be the canonical form of B.

DEFINITION 7. The *canonical form* of the matrix B is a matrix C which is row equivalent to B, with elements satisfying, for each $i = 1, \cdots, n$,

(a) $c_{ii} = 0$ or $c_{ii} = 1$;

(b) if $c_{ii} = 0$, then $c_{ij} = 0$ for all j and $c_{ji} = 0$ for $j \geq i$;

(c) if $c_{ii} = 1$, then $c_{ij} = 0$ for $j < i$ and $c_{ji} = 0$ for $j \neq i$.

Lemma 2 can be used to prove

LEMMA 4. *If $\varphi_Y(t) = \psi_Y(tB') + f(t)$, then $\psi_Y(t) = \psi_Y(tF') + \psi_Y(t(I - F)')$, where*

(i) *F is idempotent and row-equivalent to B*

(ii) *$\exp\{-\psi_Y[t(I - F)']\}$ is the characteristic function of a multinormal random variable.*

From the preceding lemma and the definition of nonnormality one easily obtains

LEMMA 5. *$nn(Y) = s$ if and only if s is the minimum rank of all matrices A such that*

$$\psi_Y(t) = \psi_Y(tA') + f(t).$$

Lemma 5 hence furnishes an alternate definition of nonnormality.

PROOF OF THEOREM 3. If $B \varepsilon \vartheta$, then $r(B) = s$. Let Y be any random variable such that $L^*(\vartheta, F_Y)$, and let $t = nn(Y)$. Then

$$(1) \qquad Y = XB + U.$$

(a) From the relation L^*, it follows that

$$(2) \qquad \psi_Y(t) = \psi_X(tB') + \psi_U(t).$$

Hence by Lemma 3, $\psi_Y(t) = \psi_Y(tC') + f(t)$, where C is idempotent and row-equivalent to B. Therefore, by Lemma 5, $t \leqq s$.

(b) Assume θ is identifiable (L^*). By Lemma 5 there exists a matrix A of rank t such that $\psi_Y(t) = \psi_Y(tA') + f(t)$. Hence there exists a matrix F having the properties enumerated in Lemma 4. But from (2) above, it follows that

$$\psi_Y(tF') = \psi_X(t(BF)') + \psi_U(tF').$$

Therefore, letting $B^* = BF$ and $U^* = UF + Y(I - F)$, one obtains that X and U^* are independent and U^* has a multinormal distribution. Further,

$$Y = XB^* + U^*,$$

so that $L^*(\vartheta^*, F_Y)$ holds, where ϑ^* is the set having B^* as an element. Now $r(B^*) \leqq r(F) = t$. But since θ is identifiable, $\mathcal{S}(\vartheta) \subset \mathcal{S}(\vartheta^*)$ so that $r(B) \leqq r(B^*)$. Whence $s \leqq t$, and part (a) then implies $s = t$, that is, the condition of the theorem is necessary.

(c) Assume θ is not identifiable. In view of part (a) it is required to show that $t < s$. By hypothesis, there exist random variables X^* and U^* and a matrix $B^* \varepsilon \vartheta^*$, such that $L^*(\vartheta^*, F_Y)$,

$$(3) \qquad Y = X^*B^* + U^*$$

and $\mathcal{S}(\vartheta) \not\subset \mathcal{S}(\vartheta^*)$. Equations (1) and (3) and Lemma 3 imply

$$(4) \qquad \psi_Y(t) = \psi_Y(tC') + f_1(t) = \psi_Y(tC^{*\prime}) + f_2(t),$$

where C and C^* are idempotent and respectively row-equivalent to B and B^* and have s and s^* rows which are not identically zero. There exists a nonsingular transformation P which reduces C^* to a diagonal matrix $D^* = C^*P$ having only ones and zeros on its main diagonal. Let $A = CP$, then (4) yields

$$\psi_Y(tA') = \psi_Y(tD^*) + f(t), \qquad \psi_Y(tD^*A') = \psi_Y(tD^*) + f(t),$$

$$(5) \qquad \psi_Y(tA') = \psi_Y(tD^*A') + f(t).$$

Equation (5) will be analyzed in three cases.

CASE I, $r(D^*A') < s$. Let $G' = P'^{-1}D^*A'$. Then $r(G) < s$, and

$$\psi_Y(tC') = \psi_Y(tG') + f(t).$$

This, together with equation (4) and Lemma 5, implies $t \leqq r(G)$, so that $t < s$.

CASE II, $r(D^*A') = s$, and there exists a diagonal matrix D having only ones and zeros on the main diagonal such that

$$r(D) = s, \qquad r(DA') = s, \qquad r(DD^*) < s.$$

Substitution of $\sigma = tD$ for t in (5) gives

$$\psi_Y(\sigma DA') = \psi_Y(\sigma DD^*A') + f(\sigma).$$

The vector $\tau = \sigma DA'$ has exactly s components which are not identically zero. Since $r(DA') = s$, there exists a matrix α such that $r(\alpha) = s$ and $\sigma = \tau\alpha$. Hence

$$\psi_Y(\tau) = \psi_Y(\tau\alpha DD^*A') + f(\tau).$$

Since tA' has the same nonvanishing components as τ,

$$\psi_Y(tC') = \psi_Y(tH') + f(t)$$

where $H' = P'^{-1}A'\alpha DD^*A'$. This, together with equation (4) and Lemma 5, implies $t \leqq r(H)$, and since $r(G) \leqq r(DD^*) < s$, it follows $t < s$.

CASE III, $r(D^*A') = s$, and for every diagonal matrix D having ones and zeros on the main diagonal, $r(DD^*) = s$ whenever $r(D) = s$ and $r(DA') = s$. Let a_{i_j} for $j = 1, \cdots, m$ be the row vectors of A' which are not identically zero. Then by Lemma 1 each row a_{i_j} is included in some s-rowed minor of A' of rank s. That is, there exists a diagonal matrix D_j such that $r(D_jA') = s$ and $r(D_j) = s$ with elements $d_{i_ji_j} = 1$ for $j = 1, \cdots, m$. Since $r(D_jD^*) = s$, by hypothesis, then $d^*_{i_ji_j} = 1$ for $j = 1, \cdots, m$. Hence, it follows that $AD^* = A$. Since D^* is idempotent, then $\mathcal{S}^*(A) \subset \mathcal{S}^*(D^*)$. Here the notation $\mathcal{S}^*(A)$ denotes the space spanned by the row vectors of A. It then follows that $\mathcal{S}^*(C) \subset \mathcal{S}^*(C^*)$, and hence $\mathcal{S}(\vartheta) \subset \mathcal{S}(\vartheta^*)$, contradicting the hypothesis that θ is not identifiable (L^*). Case III is therefore impossible.

This completes the proof of sufficiency for Theorem 3, as these three cases exhaust the possible situations arising from (5). The corollaries below are easy consequences of Theorem 3.

COROLLARY 1. *Denoting XB by S, θ is not identifiable (L^*) if and only if $S = ZG + V$, where Z and V are independent, V has a multinormal distribution, and $r(G) < r(B)$.*

COROLLARY 2. *If $X = ZG + V$ and $r(G) < r(B)$, then θ is not identifiable (L^*).*

COROLLARY 3. *If θ is identifiable (L^*), then the nonnormality of X is not less than s.*

The expression of Theorem 3 in terms of the random variables X is more natural if the problem is reformulated in an equivalent way [8]. Let Y_n denote

a vector with n components and B_{sn} a matrix with s rows and n columns, $s \leqq n$. Let $\mathfrak{X}_s$ denote a set of s-dimensional distribution functions.

THEOREM 4. *When the relation L^{**} is characterized by the equation $Y_n = X_s B_{sn} + U_n$ with $r(B_{sn}) = s$, then θ is identifiable (L^{**}) if and only if $nn(\mathfrak{X}_s) = s$.*

Taking $n = 2$ and $s = 1$, one obtains the result of Reiersøl [7]. Exactly similar results are obtained if the relation R is taken to be equality rather than inclusion. Again, if Θ_s were chosen as the set where elements are all the sets of row-equivalent matrices, so that $\Theta_s = \Theta$, then one would have:

THEOREM 5. *θ is identifiable (L^*) if no linear combination of the components of X is normally distributed.*

5. Estimation of linear structures. In this section, an estimate is constructed which converges with probability one to the linear structure. An infinite sequence of vector random variables $(Y_1, Y_2, \cdots)$ is considered. No assumption whatever is made concerning the existence of moments of Y_i. Each Y_i satisfies definitions 2 and 3; furthermore, Y_i and Y_j are independent if $i \neq j$. It is assumed that s is known. For every N, let $Z_N = (Y_1, \cdots, Y_N)$. A function $T_N(Z_N)$ will be constructed such that $P\{T_N(Z_N) \rightarrow \mathcal{S}(B) \text{ as } N \rightarrow \infty\} = 1$, where T_N is a linear vector space and the convergence $T_N(Z_N) \rightarrow \mathcal{S}(B)$ is defined in

DEFINITION 8. If C_N and C are linear vector spaces, then $C_N \rightarrow C$ as $N \rightarrow \infty$, provided $\Delta_N \rightarrow 0$ as $N \rightarrow \infty$, where $\Delta_N = \max_k \min_r |k - r|$ for all unit vectors k in C_N, and all unit vectors r in C. The quantity Δ_N will be called the *distance* between the sets C_N and C.

Hence if C_N and C are linear vector spaces and C_N a random variable, then C_N converges almost surely to C if $P\{\Delta_N \rightarrow 0\} = 1$. A unit vector is here a vector of length one.

DEFINITION 9. A matrix B is *related* to a random variable Y if $B \,\varepsilon\, \vartheta$ for some ϑ such that $\vartheta \,\varepsilon\, \Theta_s$ and $\vartheta \,\varepsilon\, \gamma(F)$ (cf. Definition 1).

From part (c) of Theorem 3 one obtains:

LEMMA 6. *If $\psi_Y(t) = \psi_Y(tC_i') + f_i(t)$ for $i = 1, 2$, where C_i is idempotent with rank S_i, then*

(i) *either $nn(Y) < s$, or else $\mathcal{S}(C_1) \subset \mathcal{S}(C_2)$, and*

(ii) *either $nn(Y) < s_2$ or else $\mathcal{S}(C_2) \subset \mathcal{S}(C_1)$.*

LEMMA 7. *If θ is identifiable (L^*), if B is related to Y, and if G is idempotent, then $\psi_Y(t) = \psi_Y(tG') + f(t)$ if and only if $\mathcal{S}(B) \subset \mathcal{S}(C)$.*

PROOF. Suppose $\mathcal{S}(B) \subset \mathcal{S}(G)$ and C is an idempotent matrix row-equivalent to B. Then $\psi_Y(t) = \psi_Y(tC') + f(t)$, and $r(C) = nn(Y)$. Since $CG = C$, then $\psi_Y(tG') = \psi_Y(tC') + f(t)$ and $\psi_Y(t) = \psi_Y(tG') + f(t)$.

Conversely, suppose $\psi_Y(t) = \psi_Y(tG') + f(t)$. Then, since $nn(Y) = r(C)$, Lemma 6 implies $\mathcal{S}(C) \subset \mathcal{S}(G)$.

Lemmas 4 and 7 imply

LEMMA 8. *If θ is identifiable (L^*) and B is related to Y, then, for any idempotent G such that $\mathcal{S}(B) \subset \mathcal{S}(G)$, there exists F idempotent and row-equivalent to G such that $\psi_Y(t) = \psi_Y(tF') + \psi_Y(t(I - F)')$, and $Y(I - F)$ has a multinomial distribution.*

LEMMA 9. *If F is idempotent with rank $n - 1$, then $F = I - r'a/ra'$ for unique row vectors r and a.*

From Lemmas 8 and 9, it follows that $\psi_Y(t) = \psi_Y(tF') + \frac{1}{2}\sigma^2(ta')^2$ if and only if $\mathcal{S}(B) \subset \mathcal{S}(F)$, where F is chosen as in Lemma 9. This property is made the basis of a criterion to determine $\mathcal{S}(B)$. Letting

$$L(t) = L(t;\quad a, r, \alpha) = \varphi_Y(t) - \varphi_Y(tF')\alpha^{-(ta')^2} \qquad \text{where } \alpha = \exp[-\tfrac{1}{2}\sigma^2],$$

it follows that $L(t) =_t 0$ if and only if $\mathcal{S}(B) \subset \mathcal{S}(F)$ and α is suitably chosen. Define $G(r) = \min_{a,\alpha} \int L(t)L(-t)\, d\lambda(t)$, where $\lambda(t)$ is a strictly increasing bounded function and the integration is taken over the entire space. Then $G(r) = 0$ if and only if r is orthogonal to $\mathcal{S}(B)$. Thus if F_Y were known, an investigation of the zeros of $G(r)$ would yield explicit knowledge of $\mathcal{S}(B)$.

Determining a random variable which converges almost surely to G will enable the desired estimate to be constructed. To this end the sample characteristic function is defined by

$$\varphi_N(t; Z_N) = \frac{1}{N}\sum_{k=1}^{N} e^{[-ity_k']}.$$

Then $G_N(r;\quad Z_N)$ is defined by replacing $\varphi_Y(t)$ by $\varphi_N(t;\quad Z_N)$ in the definition of $G(r)$. The space C is complementary to $S = \mathcal{S}(B)$, that is the space spanned by the unit vectors r for which $G(r) = 0$.

The estimate $T_N(Z_N)$ is defined to be the linear space orthogonal to the linear vector space C_N spanned by the vectors $k_1, k_2, \cdots, k_{n-s}$. The vectors k_j are defined by the following construction.

(i) k_1 is any unit vector for which $G_N(k_1;\quad Z_N) = \min_r G_N(r;\quad Z_N)$.

(ii) for $j = 2, \cdots, n - s$, k_j is any unit vector such that $G_N(k_j;\quad Z_N) = \min_{r \epsilon O_j} G_N(r;\quad Z_N)$, where O_j is the linear space orthogonal to $k_1, \cdots, k_{j-1}$.

The proof that the estimate converges almost surely is based on the following lemma which is a corollary of a theorem of Rubin [10].

LEMMA 10. *For any finite cell T of Euclidean n-dimensional space,*

$$P\{\lim_{N\to\infty} \varphi_N(t; Z_N) = \varphi_Y(t) \quad \textit{uniformly for} \quad t \,\epsilon\, T\} = 1.$$

Taking F as in Lemma 9 and since α^{-u^2} is bounded for $\alpha \,\epsilon\, [0, 1]$ and u real, then

$$P\{\lim_{N\to\infty} L_N(t) = L(t) \quad \text{uniformly for} \quad t \,\epsilon\, T \quad \text{and} \quad r, a, \alpha\} = 1.$$

Here $L_N(t)$ is defined by replacing $\varphi_Y(t)$ by $\varphi_N(t;\quad Z_N)$ in the definition of $L(t)$. From this, since r, a, α are on compact sets, it follows that

LEMMA 11. $P\{\lim_{N\to\infty} G_N(r;\quad Z_N) = G(r)$ *uniformly in* $r\} = 1$.

LEMMA 12. *If Δ_N is the distance between C_N and C, then $P\{\Delta_N \to 0$ as $N \to \infty\} = 1$, provided θ is identifiable.*

PROOF. For any $\tau > 0$, let C_τ be the set of unit vectors k such that $\min_{r \varepsilon C} |k - r| < \tau$, where r is a unit vector. For any $\eta > 0$, take

$$\tau = \eta n^{-1/2}, \qquad \xi = \min G(r) \text{ with } r \text{ not a member of } C_\tau \qquad \epsilon < \xi/2.$$

Then, there exists N_η such that for every r and all $N > N_\eta$, $| G_N(r; \; Z_N) - G(r) | < \epsilon$ with probability one and, hence, both

$$\min G_N(r; \; Z_N) \text{ with } r \text{ not a member of } C_\tau \geqq \xi - \epsilon > \xi/2$$

$$\min G_N(r; Z_N) \text{ with } r \text{ not a member of } C_\tau < \epsilon < \xi/2.$$

Therefore, if k satisfies $\min_r G_N(r; \; Z_N) = G_N(k; \; Z_N)$, it follows that $k \varepsilon C_\tau$, and hence $k_1 \varepsilon C_\tau$.

It can be shown similarly that, if $n - s \geqq 2$, then $k_2 \varepsilon C_\tau$, since in this case there must be a unit vector r such that $r \varepsilon k$ and $r \varepsilon C$. Likewise it can be shown by induction that $k_j \varepsilon C$ for $j = 1, \cdots, n - s$.

Let k be any unit vector in C_N. Then $k = \sum_{j=1}^{n-s} d_j k_j$ and $kk' = 1$ implies $\sum_{j=1}^{n-s} d_j^2 = 1$. Since $k_j \varepsilon C_\tau$ there are vectors $r_j \varepsilon C$ such that $|k_j - r_j| < \tau$ for $j = 1, \cdots, n - s$. Then

$$| k - 4 | < \tau \sum_{j=1}^{n-s} | d_j | \leqq \tau\sqrt{n - s} < \eta.$$

Hence, for any η there exists N_η such that $C_N \subset C_\eta$, provided $N > N_\eta$, and therefore $\Delta_N < \eta$ with probability one.

The following lemma is straightforward.

LEMMA 13. *C_N converges to C if and only if S_N converges to S, where C_N and C are the complements of S_N and S, respectively.*

Lemmas 12 and 13 then imply

THEOREM 6. *If θ is identifiable (L^*), then the estimate $T_N(Z_N)$ converges almost surely to $\mathcal{S}(B)$.*

ACKNOWLEDGMENT. I wish to acknowledge my debt to Professor J. Neyman for suggesting this problem and for his constant encouragement and stimulation during the development of this paper.

REFERENCES

[1] J. NEYMAN, "Existence of consistent estimates of the directional parameter in a linear structural relation between two variables," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 497–512.

[2] TJALLING C. KOOPMANS, "Identification problems in economic model construction," *Econometrica*, Vol. 17 (1949), pp. 125–144.

[3] KARL PEARSON, "On lines and planes of closest fit to systems of points in space," *Philos. Mag.* Series 6, Vol. 2 (1901), pp. 559–572.

[4] GODFREY H. THOMSON, "A hierarchy without a general factor," *British Journal of Psych.*, Vol. 8 (1916), pp. 271–281.

[5] J. NEYMAN, "Remarks on a paper by E. C. Rhodes," *Jour. Roy. Stat. Soc.*, Vol. 100 (1939), pp. 50–57.

[6] ABRAHAM WALD, "The fitting of straight lines if both variables are subject to error," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 284–300.

[7] OLAF REIERSØL, "Identifiability of a linear relation between variables which are subject to error," *Econometrica*, Vol. 18 (1950), pp. 375–389.

[8] T. A. JEEVES, "Identifiability and almost sure estimability of linear structures in n-dimensions," University of California, 1952, unpublished thesis.

[9] T. C. KOOPMANS AND O. REIERSØL, "The identification of structural characteristics," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 165–181.

[10] HERMAN RUBIN, "Uniform convergence of random functions in the applications to statistics," Technical Report No. 7, Contract No. N6-onr-251 Task III (1951), Department of Statistics, Stanford University.

ASYMPTOTIC BEHAVIOR OF SOME RANK TESTS FOR ANALYSIS OF VARIANCE[1]

By Fred C. Andrews[2]

Stanford University

1. Summary. The H test and the median test have been proposed for testing the hypothesis of the equality of c probability distributions. Assuming translation-type alternatives, we find the limiting distributions of the H and median test statistics. These results are used to derive general formulas for the asymptotic relative efficiencies of these tests with respect to one another and to the classical F test. A short discussion of the computation of approximate power functions of these tests is also included.

2. Introduction. A few tests of a non-parametric nature have been proposed for the problem of testing the equality of c probability distributions, the so called c-sample problem. Tests for the two-sample problem have been proposed by Wilcoxon [22], Mann and Whitney [11], J. Westenberg, [21], and Mood and Brown [12], among others. Consistency and power properties of some of these tests have been discussed by van Dantzig [3], Lehmann [8], [9], and Mood [13], among others.

The H test proposed by Wallis and Kruskal [20] is a direct generalization of the two-sided Wilcoxon test discussed in detail by Mann and Whitney [11]. The H test is similar to a classical F test, with ranks replacing the original observations. The Mood-Brown median test [12] makes use of the construction of a 2-by-c table and the resulting large sample theory thereof. Pitman [16] derives the general formula for the asymptotic relative efficiency of the Wilcoxon test with respect to the ordinary t test, when quite general translation-type alternative hypotheses are assumed. Mood [13] assumes only normal alternative hypotheses and computes the asymptotic relative efficiencies, with respect to the t test, to be $3/\pi$ for the Wilcoxon test and $2/\pi$ for the median test; the former is a special case of the Pitnam result.

After setting up suitable alternative hypotheses and finding the limiting distributions of the relevant statistics, we find general formulas for the asymptotic relative efficiencies in the c-sample case for translation alternatives but almost arbitrary distributions. These formulas do not in general depend on c.

Mood [12] and Kruskal [7] derive the limiting distributions of their respective statistics in the case of the hypothesis of equal distributions. The methods used here to derive the distributions under the alternative hypothesis duplicate their results.

Received 11/4/53.

[1] This work was performed under a contract with the Office of Naval Research.

[2] Now at the University of Nebraska.

A rather complete bibliography on nonparametric c-sample tests and a discussion of the rationale for applying them is given by Wallis and Kruskal [20].

The c-sample problem may be expressed formally as follows. Let $\{X_{ij}\}$ for $i = 1, 2, \cdots, c$, and $j = 1, 2, \cdots, n_i$ be a set of independent random variables. The probability distribution function of X_{ij} is denoted by F_i, so that $F_i(x)$ is the probability of the event $[X_{ij} \leqq x]$. The set of admissible hypotheses designates that each F_i belongs to some class of distribution functions Ω. The hypothesis to be tested, say K_0, specifies that F_i is an element of Ω for each i and that furthermore $F_1(x) = F_2(x) = \cdots = F_c(x)$ for all real x. Alternative to K_0 is the hypothesis that each F_i belongs to Ω but that K_0 does not hold. To avoid the problem of ties, it is assumed throughout that the class Ω is the class of continuous distribution functions.

So as to pay particular attention to translation-type alternatives, the class of admissible hypotheses will be limited to include only those hypotheses which state that $F_i(x) = F(x + \epsilon_i)$ for all $i = 1, 2, \cdots, c$, for some arbitrary choice of F in the class Ω and real numbers $\epsilon_1, \cdots, \epsilon_c$. It is seen immediately that specifying all $\epsilon_i = 0$ yields hypothesis K_0, while if $\epsilon_i \neq \epsilon_j$ for some pair (i, j) then an alternative to K_0 is given.

The H test is based on the statistic

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{c} n_i \left(\bar{R}_i - \frac{N+1}{2}\right)^2, \tag{1}$$

when $\bar{R}_i$ is the average rank of the members of the ith sample obtained after ranking all of the $N = \sum n_i$ observations. The test consists in rejecting K_0 at a significance level α if H exceeds some predetermined number h_α. Kruskal [7] proves that if K_0 is true, the statistic H has a limiting chi square distribution with $c - 1$ degrees of freedom as all $n_i \to \infty$ simultaneously. This provides the user of this H test with a large sample approximation of the value of h_α for any $0 < \alpha < 1$.

The Mood-Brown median test is based on the statistic

$$M = \frac{N(N-1)}{b(N-b)} \sum_{i=1}^{c} \frac{1}{n_i} \left(m_i - \frac{bm_i}{N}\right)^2 \tag{2}$$

where $N = \sum n_i$, and $b = \frac{1}{2}(N-1)$ when N is odd, and $b = \frac{1}{2}N$ when N is even, while m_i denotes the number of observations in the ith sample which are less than the median of all of the observations. Mood proves that whenever K_0 is true, the statistic M has a limiting chi square distribution with $c - 1$ degrees of freedom as all $n_i \to \infty$ simultaneously. The median test is then to reject K_0 at the level of significance α whenever M exceeds a number m_α. Use of the limiting distribution allows one to determine an approximate value for m_α for large samples.

Because both the H test and the median test are consistent against translation alternatives, the distributions of H and M will be studied, in following sections, assuming a sequence of admissible alternative hypotheses K_n for $n = 1, 2, \cdots$.

The hypothesis K_n specifies, for each $i = 1, 2, \cdots, c$, that $F_i(x) = F(x + \theta_i/\sqrt{n})$, with $F \varepsilon \Omega$ but not specified further, and for some pair (i, j) that $\theta_i \neq \theta_j$. The letter n will be used to index a sequence of situations in which K_n is the true hypothesis. Limiting probability distributions will then be found as $n \to \infty$. The problem will be so formulated that N will be proportional to n. For large n the hypothesis K_n is "near" K_0, so that this type of limit process provides a way of studying the effect of small translations on these tests.

The notation $\chi_r^2(\lambda)$ will denote the possibly noncentral chi square distribution with degrees of freedom r and noncentral parameter λ. Thus $\chi_r^2(\lambda)$ is the probability distribution of the sum of r squares of independent normal random variables whose variances are all unity and whose sum of squared expectations is denoted by λ. For $\lambda = 0$ we see that $\chi_r^2(0)$ is the ordinary chi square distribution. The $\chi_r^2(\lambda)$ distribution has been studied and used by Fisher [4], Tang [19], and Patnaik [15], among others. A partial tabulation of some percentage points of $\chi_r^2(\lambda)$ is given by Fix [5].

3. The limiting distribution of H under hypothesis K_n. The purpose of this section is to prove

THEOREM 3.1. *For each index n assume that $n_\alpha = s_\alpha n$, with s_α a positive integer, and the truth of hypothesis K_n. If for any real number t,*

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} \sqrt{n}\,\{F(x + t/\sqrt{n}) - F(x)\}\, dF(x)$$

exists finite, then, for $n \to \infty$, the limiting distribution of the statistic H is $\chi_{c-1}^2(\lambda'')$, where

$$\lambda'' = \left[12\left(\sum_{j=1}^{c} s_j\right)^{-2}\right] \sum_{\alpha=1}^{c} s_\alpha \cdot \left\{\sum_{i=1}^{c} s_i \lim_{n\to\infty} \int_{-\infty}^{+\infty} \sqrt{n}\left[F\left(x + \frac{\theta_i - \theta_\alpha}{\sqrt{n}}\right) - F(x)\right] dF(x)\right\}^2. \tag{3}$$

From (1) and definitions (5), (6), and (9) below one can write

$$H = \left[12\Big/\left(\sum_{i=1}^{c} s_i\right)\left(\sum_{i=1}^{c} s_i + \frac{1}{n}\right)\right] \sum_{\alpha=1}^{c} s_\alpha^3 \left[\sqrt{n}\left(U'^\alpha - \tfrac{1}{2} \sum_{i=1\neq\alpha}^{c} \frac{s_i}{s_\alpha}\right)\right]^2. \tag{4}$$

The proof of Theorem 3.1 then quite naturally depends upon showing that the random variables

$$\sqrt{n}\left(U'^\alpha - \tfrac{1}{2} \sum_{i=1\neq\alpha}^{c} \frac{s_i}{s_\alpha}\right), \qquad \alpha = 1, 2, \cdots, c,$$

have a certain joint limiting normal distribution as $n \to \infty$. The methods used in the proof are mainly adaptations of results of Hoeffding [6] and Lehmann [8].

We begin the proof by defining the functions h^α by

$$h^\alpha(y_1, \cdots, y_\alpha, \cdots, y_c) = \sum_{\beta=1}^{c} \frac{s_\beta}{s_\alpha} \delta(y_\beta, y_\alpha), \tag{5}$$

with the convention that $\delta(y_\beta, y_\alpha) = 1$ whenever $y_\beta < y_\alpha$ and is otherwise zero. Throughout this discussion α will range over the integers 1, 2, $\cdots$, c. Recalling that $n_i = s_i n$, we construct for $k = 1, 2, \cdots, n$ the random vectors

$$\begin{aligned} X_k = (X_{1\ (k-1)s_1+1}, X_{1\ (k-1)s_1+2}, \cdots, X_{1\ ks_1}; \\ X_{2\ (k-1)s_2+1}, \cdots, X_{2\ ks_2}; \quad \cdots \quad ; X_{c\ (k-1)s_c+1}, \cdots, X_{c\ ks_c}) \end{aligned} \tag{6}$$

and the random variables φ^α, U^α, and U'^α, defined by

$$\varphi^\alpha(X_1, \cdots, X_c) = \frac{1}{(s_1 s_2 \cdots s_c)(c!)} \sum h^\alpha(X_{1j_1}, \cdots, X_{cj_c}), \tag{7}$$

with the summation extending over all indices $(j_1, \cdots, j_c)$ in such a manner the arguments of a single h^α are components of distinct vectors;

$$U^\alpha = \sum \varphi^\alpha(X_{\beta_1}, X_{\beta_2}, \cdots, X_{\beta_c}) \Big/ \binom{n}{c} \tag{8}$$

where the summation extends over all indices $1 \leqq \beta_1 < \beta_2 < \cdots < \beta_c \leqq n$; and

$$U'^\alpha = \frac{1}{n_1 n_2 \cdots n_c} \sum_{j_1=1}^{n_1} \cdots \sum_{j_c=1}^{n_c} h^\alpha(X_{1j_1}, X_{2j_2}, \cdots, X_{cj_c}). \tag{9}$$

Then U'^α is recognized as the average of all kinds of h^α terms while U^α is an average of only those h^α terms in which the arguments of a given h^α are each elements of a different vector. Setting j^α equal to the sum of all h^α terms appearing in U'^α but not in U^α, we have

$$U'^\alpha = \frac{1}{n_1 n_2 \cdots n_c} \left\{ \binom{n}{c} U^\alpha + J^\alpha \right\}.$$

Let

$$D^\alpha = U'^\alpha - U^\alpha = \frac{1}{n_1 n_2 \cdots n_c} \left\{ \left[\binom{n}{c} c! - n_1 n_2 \cdots n_c \right] U^\alpha + J^\alpha \right\}.$$

Adopting a method of proof given by Lehmann ([8], p. 168), we use the inequality $(\sum_1^k a_i)^2 \leqq k \sum_1^k a_i^2$ for real numbers a_i, and the fact that

$$E\{h^\alpha(X_{1j_1}, X_{2j_2}, \cdots, X_{cj_c})\}^2 \leqq \left(\sum_{\beta=1}^{c} \frac{s_\beta}{s_\alpha} \right)^2.$$

Thus we establish that

$$E(\sqrt{n}\, D^\alpha)^2 \leqq 4 \left(\sum_{\beta=1}^{c} \frac{s_\beta}{s_\alpha} \right)^2 \left[\sqrt{n} \left(1 - \frac{1}{n_1 n_2 \cdots n_c} \binom{n}{c} \right) \right]^2 \to 0 \quad \text{as } n \to \infty.$$

With the notation $W^\alpha = \sqrt{n}\,(U'^\alpha - EU'^\alpha)$ and $Z^\alpha = \sqrt{n}\,(U^\alpha - EU^\alpha)$,

$$E(W^\alpha - Z^\alpha)^2 = E(\sqrt{n}\,D^\alpha)^2 \to 0 \qquad \text{as } n \to \infty.$$

Applying Lemma 7.1 of Hoeffding ([6], p. 305), we have

LEMMA 3.1. *If either of the random vectors* $W = (W^1, W^2, \cdots, W^c)$ *or* $Z = (Z^1, Z^2, \cdots, Z^c)$ *has a limiting probability distribution as* $n \to \infty$, *then the other random vector has this same limiting distribution as* $n \to \infty$.

The next step in the proof is to compare the random vector Z with the random vector $Y = (Y^1, Y^2, \cdots, Y^c)$, whose components are defined by $Y = (c/\sqrt{n})\sum_1^n \psi_1^\alpha(X_i)$, with

$$\psi_j^\alpha(x_1, x_2, \cdots, x_j) = E\varphi^\alpha(x_1, x_2, \cdots, x_j, X_{j+1}, X_{j+2}, \cdots, X_c) - E\varphi^\alpha(X_1, X_2, \cdots, X_c).$$

The functions $\psi_j^\alpha(x_1, x_2, \cdots, x_j)$ are the same as those defined by Hoeffding [6] except that they are applied to this special problem. Now Hoeffding ([6], p. 299, (5.13)) has shown that

$$E(Z^\alpha)^2 = n\sigma^2(U^\alpha) = cn\binom{n}{c}^{-1}\binom{n-c}{c-1}a_1^\alpha + R_{nc}^\alpha,$$

$$R_{nc}^\alpha = n\binom{n}{c}^{-1}\sum_{j=2}^{c}\binom{c}{j}\binom{n-c}{c-j}a_j^\alpha, \qquad a_j^\alpha = E\{\psi_j^\alpha(X_1, X_2, \cdots, X_j)\}^2.$$

By expanding binomial coefficients we calculate

$$R_{nc}^\alpha = \sum_{j=2}^{c}(j!)\binom{c}{j}^2\prod_{k=1}^{c-j}\left[1 - \frac{c-1}{n-k}\right]\sum_{l=1}^{j-1}[n - c + j - l]^{-1}a_j^\alpha;$$

however, $a_j^\alpha \leqq 4(\sum_{\beta=1}^{c} s_\beta/s_\alpha)^2$ for all α, j so that $R_{nc}^\alpha \to 0$ as $n \to \infty$. Referring to Hoeffding ([6], p. 308, (7.10) and (7.12)), we find that

$$E(Y^\alpha)^2 = E(Y^\alpha Z^\alpha) = c^2a_1^\alpha.$$

Substitution yields

$$E(Y^\alpha - Z^\alpha)^2 = E(Y^\alpha)^2 + E(Z^\alpha)^2 - 2E(Z^\alpha Y^\alpha) = R_{nc}^\alpha + \left[cn\binom{n}{c}^{-1}\binom{n-c}{c-1} - c^2\right]a_1^\alpha \to 0 \qquad \text{as } n \to \infty.$$

Another application of Lemma 7.1 of Hoeffding ([6], p. 305), produces

LEMMA 3.2. *If either of the random vectors* Z *or* Y *has a limiting probability distribution as* $n \to \infty$, *then the other random vector has this same limiting distribution as* $n \to \infty$.

It now remains to find the limiting distribution of Y. Each Y^α is a sum of independent and identically distributed random variables,

$$Y^\alpha = \frac{c}{\sqrt{n}}\sum_{j=1}^{n}\psi_1^\alpha(X_j); \qquad E(Y^\alpha) = 0.$$

Also, $\psi_1^\alpha(X_j) \leqq 2(\sum_{\beta=1}^c s_\beta/s_\alpha)$ with probability one. Adopting the notation for the column vectors $s = (s_1, \cdots, s_c)'$ a vector of real numbers, and $\psi_1 = (\psi_1^1, \psi_1^2, \cdots, \psi_1^c)'$, the characteristic function of Y is expressed as

$$f_n(s) = E(e^{is'Y}) = E(\exp\{ics'\psi_1/\sqrt{n}\})^n,$$

because of independence. Taking logarithms, expanding the real and imaginary parts of the exponential in finite Taylor series, using the almost sure boundedness of $\psi_1(X_j)$, noting that $E[\psi_1(X_j)] = 0$, and finally expanding the logarithm in a finite Taylor series, produces the usual type of result that

$$\log f_n(s) = -\tfrac{1}{2}c^2s'[E(\psi_1\psi_1')]s + O(n^{-1/2}) \qquad \text{as } n \to \infty$$

for any fixed real vector s. From the continuity theorem for characteristic functions ([2], p. 96), we conclude

LEMMA 3.3. *The random vector Y has a limiting normal distribution with $E(Y)$ the zero vector and variance-covariance matrix $\Sigma = \lim_{n\to\infty} c^2E(\psi_1\psi_1')$.*

Adopting the notation

$$A_{\beta\alpha} = \frac{1}{s_\alpha}\sum_{j=1}^{s_\alpha}\left[F_\beta(X_{\alpha j}) - \int F_\beta(x)\, dF_\alpha(x)\right] - \frac{1}{s_\beta}\sum_{j=1}^{s_\beta}\left[F_\alpha(X_{\beta j}) - \int F_\alpha(x)\, dF_\beta(x)\right],$$

we can recognize $\psi_1^\alpha(X_1) = (1/c)\sum_{\beta=1}^c (s_\beta/s_\alpha)A_{\beta\alpha}$. A lengthy computation and an application of the Lebesgue bounded convergence theorem, in view of the boundedness of each F_j and $\lim_{n\to\infty} F_j(x) = F(x)$, yields the result that

$$\Sigma = \lim_{n\to\infty} c^2E(\psi_1\psi_1') = \tfrac{1}{12}\left[\frac{1}{s_\alpha}\left(\sum_{j=1}^c \frac{s_j}{s_\beta}\right)\left(\delta_{\alpha\beta}\sum_{j=1}^c \frac{s_j}{s_\beta} - 1\right)\right].$$

Combining the previous three lemmas produces

LEMMA 3.4. *If for each index n the hypothesis K_n is valid and W^α denotes the random variable $\sqrt{n}\,(U'^\alpha - EU'^\alpha)$, then the random vector $W = (W^1, W^2, \cdots, W^c)$* has a limiting normal distribution with zero mean vector and variance-covariance matrix Σ.

Recalling $W^\alpha = \sqrt{n}\,(U'^\alpha - E(U'^\alpha))$ and (4), and letting

$$m^\alpha = \sqrt{n}\left[E(U'^\alpha) - \tfrac{1}{2}\sum_{i=1}^c \frac{s_i}{s_\alpha}\right],$$

we write H as

$$H = \left[12\Big/\left(\sum_{i=1}^c s_i\right)\left(\sum_{i=1}^c s_i + \frac{1}{n}\right)\right]\sum_{\alpha=1}^c s_\alpha^3(W^\alpha + m^\alpha)^2.$$

Now H will have the same limiting distribution as

$$H^* = \left[12\Big/\left(\sum_{i=1}^c s_i\right)^2\right]\sum_{\alpha=1}^c s_\alpha^3(W^\alpha + m^\alpha)^2,$$

but because $\sum n_\alpha \bar{R}_\alpha = \frac{1}{2}(\sum n_\alpha)(\sum n_\alpha + 1)$, we have $\sum s_\alpha^2 W^\alpha = O(n^{-1/2})$ as $n \to \infty$. So, except for terms of higher order,

$$H^* = \left[12 \Big/ \left(\sum_{i=1}^{c} s_i\right)^2\right] \sum_{\alpha=1}^{c} \sum_{\beta=1}^{c-1} s_\alpha^2 \left(s_\alpha \delta_{\alpha\beta} + \frac{s_\beta^2}{s_c}\right)(W^\alpha + m^\alpha)(W^\beta + m^\beta).$$

We recognize the matrix of the quadratic form H^* as the inverse of the limiting variance-covariance matrix of the random variables $W^1, W^2, \cdots, W^{c-1}$.

LEMMA 3.5. *If the vector x has a normal distribution, with mean vector μ and non-singular variance-covariance matrix Λ, then the quadratic form $x'\Lambda^{-1}x$ has a $\chi_r^2(\lambda)$ distribution, with $\lambda = \mu'\Lambda^{-1}\mu$ and r the rank of Λ.*

A proof of this lemma is given by Rao ([17], p. 57). We now calculate

$$\lim_{n\to\infty} m^\alpha = \sum_{\beta=1}^{c} \frac{s_\beta}{s_\alpha} \lim_{n\to\infty} \sqrt{n} \int_{-\infty}^{+\infty} \left[F\left(x + \frac{\theta_\beta - \theta_\alpha}{\sqrt{n}}\right) - F(x)\right] dF(x),$$

and combine Lemmas 3.4 and 3.5 with a theorem of Mann and Wald ([10], p. 223) to complete the proof of Theorem 3.1.

In many instances λ^H can easily be computed with the aid of

LEMMA 3.6. *If the distribution function F possesses a continuous derivative F' except at most on a set S where $\int_S dF(x) = 0$, and if there exists a function g which bounds the difference quotient $|[F(x + \theta) - F(x)]/\theta| \leq g(x)$ for which $\int_{-\infty}^{+\infty} g(x)\, dF(x) < \infty$, then*

$$\lim_{n\to\infty} \sqrt{n} \int_{-\infty}^{+\infty} [F(x + \theta/\sqrt{n}) - F(x)]\, dF(x) = \theta \int_{-\infty}^{+\infty} F'(x)\, dF(x).$$

This lemma is proved by a direct application of the Lebesgue bounded convergence theorem and the definition of the derivative. In the event that the conditions of Lemma 3.6 are satisfied, then

$$\lambda^H = 12\left\{\int_{-\infty}^{+\infty} F'(x)\, dF(x)\right\}^2 \sum_{\alpha=1}^{c} s_\alpha(\theta_\alpha - \bar{\theta})^2,$$

$$\bar{\theta} = \sum_{\alpha=1}^{c} s_\alpha \theta_\alpha \Big/ \sum_{\alpha=1}^{c} s_\alpha .$$

4. The limiting distribution of M under hypothesis K_n. The purpose of this section is to derive the limiting distribution of the statistic M as $n \to \infty$. The result is stated in

THEOREM 4.1. *Assume for each index $n = 1, 2, \cdots$ the validity of hypothesis K_n, that F has a continuous derivative F' at its median a, and that $n_i = s_i n$, for each $i = 1, 2, \cdots, c$, with s_i a positive integer. With these assumptions the limiting distribution of M is $\chi_{c-1}^2(\lambda^M)$ with*

$$\lambda^M = 4[F'(a)]^2 \sum_{i=1}^{c} s_1(\theta_i - \bar{\theta})^2, \qquad \theta = \sum_{i=1}^{c} s_i\theta_i \Big/ \sum_{i=1}^{c} s_i . \tag{10}$$

The proof of the theorem is a generalization of a type of proof sketched by Mood [13] in his discussion of the two-sample problems. Because the two cases N odd and N even require slight differences in exposition, only the proof for N odd will be given here. A similar proof for N even could readily be constructed. In this case N odd,

$$M = \frac{4}{1 + 1/N} \sum_{i=1}^{c} n_i \left(\frac{m_i}{n_i} - \tfrac{1}{2} + \frac{1}{2N}\right)^2.$$

Defining the random variables $v_j = \sqrt{n_j}\,[(m_j/n_j) - \tfrac{1}{2}]$, permits M to be written

$$M = \frac{4}{1 + 1/N} \sum_{i=1}^{c} \left\{v_j^2 + \frac{\sqrt{n_j}}{N} v_j + \frac{n_j}{4N^2}\right\}. \tag{9}$$

Provided that we can demonstrate that v_j has a limiting distribution, since $\sqrt{n_j}/N$, n_j/N^2, and $1/N$ all converge to zero as $n \to \infty$, M will have the same limiting distribution as the statistic $4 \sum v_j^2$. The first part of proof consists in proving

LEMMA 4.1. *Assuming the hypothesis of Theorem* 4.1, *the limiting distribution of the vector* $(v_1, \cdots, v_{c-1})$ *is normal with* $E(v_j) = F'(a)\sqrt{s_j}(\theta_j - \bar{\theta})$ *and covariance matrix* A_c *given by*

$$A_c^{-1} = \left(\frac{4}{s_c}\,(s_c\delta_{ij} + \sqrt{s_i s_j})\right), \qquad i, j = 1, 2, \cdots, c - 1,$$

where δ_{ij} *is the usual Kronecker delta.*

Let $r_1, \cdots, r_c$ be a set of independent random variables each with a uniform probability distribution on the unit interval and let

$$v_j' = \sqrt{n_j}\left[\frac{m_i + r_j}{n_j} - \tfrac{1}{2}\right] = v_j + \frac{r_j}{\sqrt{n_j}}, \qquad j = 1, 2, \cdots, c.$$

The difference $v_j' - v_j$ tends to zero in probability and so, by a well known theorem ([2], p. 299) the vectors $(v_1', \cdots, v_c')$ and $(v_1, \cdots, v_c)$ possess the same limiting distribution if they have one at all. Because the v_j are discrete while the v_j' are continuous random variables, it is easier to examine the limiting distribution of $(v_1', \cdots, v_{c-1}')$.

Denoting by Z the median of all the samples combined, the probability of the joint event $m_1 = a_1$ and $m_2 = a_2$ and $\cdots$ and $m_{c-1} = a_{c-1}$ and $z_1 \leqq Z \leqq z_2$ is

$$P[m_1 = a_1, \quad \cdots, \quad m_{c-1} = a_{c-1}, \quad z_1 \leqq Z \leqq z_2]$$
$$= \sum_{i=1}^{c} \int_{z_1}^{z_2} \frac{(n_i - a_i)}{1 - F_i(z)} F_i'(z) \prod_{j=1}^{c} \binom{n_j}{a_j} \{F_j(z)\}^{a_j} \{1 - F_j(z)\}^{n_j - a_j}\, dz$$

for $\sum a_i = \tfrac{1}{2}(N - 1)$ with a_i a nonnegative integer, and is zero otherwise. Writing $m_i' = m_i + r_i$ and square brackets to indicate the "largest integer contained in," we see that the joint probability density function of the random

variables $m_1', \cdots, m_{c-1}', Z$ is

$$g(m_1', \cdots, m_{c-1}', z) = \sum_{i=1}^{c} \frac{n_i - [m_i']}{1 - F_i(z)} F'(z) \sum_{j=1}^{c} \binom{n_j}{[m_j']} \{F_j(z)\}^{[m_j']} \{1 - F_j(z)\}^{n_j - [m_j']}$$

for $\sum [m_i'] = \frac{1}{2}(N - 1)$, and otherwise zero. With the transformation

$$w = \sqrt{n}\,(Z - a); \quad v_j' = \sqrt{n_j}\{(m_j'/n_j) - \tfrac{1}{2}\}, \quad j = 1, 2, \cdots, c,$$

the probability density of $(v_1', \cdots, v_{c-1}', w)$ becomes

$$h(v_1', \cdots, v_{c-1}', w) = \sum_{i=1}^{c} \frac{[d_i](n_i n_c)^{-1/2}}{1 - F_i(a + w/\sqrt{n})} F_i'\left(a + \frac{w}{\sqrt{n}}\right)$$

$$\times \prod_{j=1}^{c} \sqrt{n_j} \binom{n_j}{[d_j]} \left\{F_j\left(a + \frac{w}{\sqrt{n}}\right)\right\}^{[d_j]} \left\{1 - F_j\left(a + \frac{w}{\sqrt{n}}\right)\right\}^{n_j - [d_j]}$$

where $d_j = \frac{1}{2}n_j + v_j'\sqrt{n_j}$ and square brackets indicate the "largest integer contained in."

Noting that $\sum_1^c \sqrt{s_i}v_i' = o(1)$ as $n \to \infty$, employing Stirling's formula for $\log n!$, and using series expansions and the continuity of F' at $x = a$, we compute

$$\lim_{n\to\infty} h(v_1', \cdots, v_{c-1}', w) = \left(\frac{1}{\sqrt{2\pi}}\right)^{c-1} \frac{2^{c-1}\sqrt{s}}{\sqrt{s_c}}$$

$$\times \exp\left\{-\tfrac{1}{2} \sum_{j=1}^{c} 4[v_j - F'(a)\sqrt{s_j}(\theta_j - \bar{\theta})]^2\right\}$$

$$\times \frac{2F'(a)\sqrt{s}}{\sqrt{2\pi}} \exp\{-2(F'(a))^2 s(\bar{\theta} - w)^2\},$$

where $s = s_1 + s_2 + \cdots + s_c$. Letting A_c denote the variance-covariance matrix of $(v_1', \cdots, v_{c-1}')$, we find

$$A_c^{-1} = \{(4/s_c)(s_c\delta_{ij} + \sqrt{s_i s_j})\}, \qquad i, j = 1, 2, \cdots, c - 1.$$

Applying a theorem of Scheffé [18] yields the result that the limiting distribution of $(v_1', \cdots, v_{c-1}', w)$ is the foregoing normal distribution. Integrating out the variable w, we obtain the desired limiting probability distribution of $(v_1', \cdots, v_{c-1}')$ and hence of $(v_1, \cdots, v_{c-1})$, which proves Lemma 4.1.

Earlier in this section it was remarked that if $(v_1, \cdots, v_{c-1})$ has a limiting distribution, then H has the same limiting distribution as

$$4\sum_{i=1}^{c} v_i^2 = \left\{\sum_{j=1}^{c-1} (s_c + s_j)v_j^2 + \sum_{j=1}^{c-1} \sum_{k=1 \neq j}^{c-1} \sqrt{s_j s_k}\, v_j v_k\right\} + \eta.$$

However, η tends to zero in probability, since $\sum \sqrt{s_i}v_i = o(1)$ is satisfied with probability one. We recognize then that, except for the term η, $4\sum v_i^2$ is equal to the quadratic form in the limiting distribution of $(v_1, \cdots, v_{c-1})$, provided that the means are shifted to zero. As in Section 3, we employ Lemma 3.5 to obtain the main Theorem 4.1 of this section.

5. Asymptotic relative efficiency. The concept of asymptotic relative efficiency of one consistent test with respect to another is due to Pitman [16]. An application and account of this method of comparing consistent tests is presented by Noether ([14], p. 241). Briefly, the idea of asymptotic relative efficiency is to choose a sequence of alternative hypotheses which vary with the sample sizes in such a manner that the powers of the two tests for this sequence of alternatives have a common limit less than one. The comparison of the two tests is then made on a sample size basis.

To be more definite, suppose that two consistent tests T and T' require N and N' observations, respectively, to attain the power β at level of significance α for testing the hypothesis K_0 against hypothesis K_n. The difference in the sample sizes N and N' results from the fact that we demand that the tests yield a common power for a given alternative K_n. The asymptotic relative efficiency of T' with respect to T is defined to be

$$\lim_{N\to\infty} N/N' = \lim_{n\to\infty} n/n' = \epsilon_{T',T}(\alpha, \beta, K_0, \{K_n\}).$$

The asymptotic relative efficiency of the median test with respect to the H test is stated in

THEOREM 5.1. *If $n_i = s_i n$ and if the distribution function F has the two properties,*

(i) *F is continuous at its median, and*

(ii) *$\lim_{n\to\infty} \sqrt{n} \int_{-\infty}^{+\infty} [F(x + \theta/\sqrt{n}) - F(x)]\, dF(x)$ exists,*

then the asymptotic relative efficiency of the median test with respect to the H test for testing the hypothesis K_0 against K_n is

$$\epsilon_{M,H} = \frac{\left(\sum_{j=1}^{c} s_j\right)^2 \cdot [F'(a)]^2 \sum_{i=1}^{c} s_i(\theta_i - \bar{\theta})^2}{3 \sum_{\alpha=1}^{c} s_\alpha \left\{\sum_{i=1}^{c} s_i \lim_{n\to\infty} \int_{-\infty}^{+\infty} \sqrt{n}\left[F\left(x + \frac{\theta_i - \theta_\alpha}{\sqrt{n}}\right) - F(x)\right] dF(x)\right\}^2}.$$

To prove Theorem 5.1, let n' and n index the sample sizes for the H test and the median test, respectively. The alternative hypothesis K_n states that $F_i(x) = F(x + \theta_i/\sqrt{n})$ and so is characterized by the numbers $\theta_i/\sqrt{n}$. If the level of significance is fixed at α and the limiting power fixed at β, then, since from Theorems 3.1 and 4.1 H has a limiting $\chi^2_{c-1}(\lambda^H)$ distribution and M has a limiting $\chi^2_{c-1}(\lambda^M)$ distribution under K_n, we must have $\lambda^H = \lambda^M$ to achieve the same limiting power for the two tests. To have the same alternatives for each test we must have $\theta_i/\sqrt{n} = \theta'_i/\sqrt{n'}$. The substitution $\theta'_i = \theta_i \sqrt{n'/n}$ in (10) along with the requirement $\lambda^H = \lambda^M$ (to guarantee equal power) yields formula (11), which proves Theorem 5.1.

COROLLARY 5.1. *If in addition to the hypothesis of Theorem 6.1, the hypothesis of Lemma 3.6 is assumed, then*

$$\epsilon_{M,H} = \tfrac{1}{3}\left[F'(a) \Big/ \int_{-\infty}^{+\infty} F'(x)\, dF(x)\right]^2.$$

Here $\epsilon_{M,H}$ does not depend upon α, β, $(\theta_1, \cdots, \theta_c)$, or c, but is a function of F only.

The comparison of the H test with respect to the ordinary analysis of variance $\mathfrak{F}$ test is contained in

THEOREM 5.2. *If the distribution function F satisfies the conditions of Lemma 3.6 and if* $\int_{-\infty}^{+\infty} x^2\, dF(x) - \left[\int_{-\infty}^{+\infty} x\, dF(x)\right]^2 = \sigma_F^2$ exists, then

$$\epsilon_{H,F} = 12\sigma_F^2\left[\int_{-\infty}^{+\infty} F'(x)\, dF(x)\right]^2.$$

The classical $\mathfrak{F}$ statistic in this instance is defined by

$$\mathfrak{F} = \frac{\frac{1}{c-1}\sum_{i=1}^{c} n_i(x_{i\cdot} - \bar{x})^2}{\left[1\Big/\sum_{k=1}^{c}(n_k - 1)\right]\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{ij} - x_{i\cdot})^2}.$$

Now Fisher [4] and Tang [19] have shown that if $F(x)$ is the normal distribution function, then under hypothesis K_n the statistic $\mathfrak{F}$ has a limiting $\chi_{c-1}^2(\chi^{\mathfrak{F}})$ distribution with $\lambda^{\mathfrak{F}} = \sum_1^c s_i[(\theta_i - \bar{\theta})/\sigma_F]^2$. However, it is a well known result of the weak Law of Large Numbers that $[1/(n-1)]\sum_{j=1}^{n}(x_{ij} - x_{1\cdot})^2 \to \sigma_F^2$ in probability as $n \to \infty$. Also the Lindeberg-Levy central limit theorem shows that $\sqrt{n_i}[x_{i\cdot} - E(x_{i\cdot})]/\sigma_F$ has a limiting $N(0, 1)$ distribution. Application of the Mann-Wald theorem used previously gives the result that under hypothesis K_n the statistic $\mathfrak{F}$ has a limiting $\chi_{c-1}^2(\lambda^{\mathfrak{F}})$ distribution whenever F satisfies the hypothesis of Theorem 5.2. A calculation similar to that for the proof of Theorem 5.1 completes the proof of Theorem 5.2.

Theorems 5.1 and 5.2 show that, depending upon F, $\epsilon_{M,H}$ can be $\gtreqless 1$, similarly for $\epsilon_{H,\mathfrak{F}}$ and $\epsilon_{M,\mathfrak{F}} = \epsilon_{M,H}\epsilon_{H,\mathfrak{F}}$. In the event that F is some normal distribution function, then $\epsilon_{M,H} = 2/3$, $\epsilon_{H,\mathfrak{F}} = 3/\pi$, and $\epsilon_{M,\mathfrak{F}} = 2/\pi$. When F is the uniform distribution function on the unit interval, $F(x) = x$ if $0 \leq x \leq 1$, then $\epsilon_{M,H} = 1/3$, $\epsilon_{H,\mathfrak{F}} = 1$, and $\epsilon_{M,\mathfrak{F}} = 1/3$.

6. Power functions. The power of a test for a given simple alternative hypothesis is the probability that the test will reject the hypothesis tested when the given alternative is true. In terms of this power definition, the power function is defined on the class of alternative hypotheses.

As we have seen in Sections 3 and 4, both the H and M statistics have limiting noncentral chi square distributions when the alternatives K_n are true for each n. In the event that Lemma 3.6 is satisfied, the noncentral parameter in each of these limiting distributions is a function of $\theta_1, \cdots, \theta_c$ only through the variable $\sum s_i(\theta_i - \bar{\theta})^2$. In fact

$$\lambda^H = 12\left[\int_{-\infty}^{+\infty} F'(x)\, dF(x)\right]^2 \sum_{i=1}^{c} s_i(\theta_i - \bar{\theta})^2, \qquad \lambda^M = 4[F'(a)]^2 \sum_{i=1}^{c} s_i(\theta_i - \bar{\theta})^2.$$

For particular choices of F the power function of each of these tests could be considered as a function of $\sum s_i(\theta_i - \bar{\theta})^2$. This type of power function approximation is discussed in Cochran's paper on the chi square test for goodness of fit [1].

The tables of Fix [5] may be employed to find approximate values for these power functions. The procedure would be as follows. Suppose that F is the uniform distribution function on the unit interval, then $\lambda^H = 12\sum s_i(\theta_i - \bar{\theta})^2$ and $\lambda^M = 4\sum s_i(\theta_i - \bar{\theta})^2$. If the approximate power is desired for the test using $n_i^0 = s_i n^0$ observations in the ith sample, when alternative $F_i(x) = F(x + \epsilon_i)$ for $i = 1, 2, \cdots, c$, is true, set $\epsilon_i = \theta_i/\sqrt{n^0}$ and compute

$$\lambda^H = 12\sum_{i=1}^{c} n_i^0(\epsilon_i - \bar{\epsilon})^2, \qquad \lambda^M = 4\sum_{i=1}^{c} n_i^0(\epsilon_i - \bar{\epsilon})^2.$$

For the given level of significance and $c - 1$ degrees of freedom, enter the Fix tables and find the approximate powers for these two tests at the given alternative. Because of the limited extent of the Fix tables, the power can be found only to the first decimal place without some sort of interpolation. In most instances, however, this accuracy should be sufficient, as it is not known how close these approximations are to the true power.

7. Acknowledgement. The author wishes to express his thanks to Professor Erich L. Lehmann for proposing this investigation and for sustained interest and helpful suggestions during its progress.

REFERENCES

[1] W. G. Cochran, "The χ^2 test of goodness of fit," *Ann. Math. Stat.* Vol. 23 (1952), pp. 315–345.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] D. van Dantzig, "On the consistency and the power of Wilcoxon's two sample test," *Nederl. Akad. Wetensch. Proc.* Vol. 54 (1951), pp. 1–8.

[4] R. A. Fisher, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc. London. Ser. A.*, Vol. 121 (1928), pp. 654–673.

[5] E. Fix, "Tables of the noncentral χ^2," *Univ. California Publ. Stat.*, Vol. 1 (1949), pp. 15–19.

[6] W. Hoeffding, "A class of statistics with asymptotically normal distribution," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 293–325.

[7] W. H. Kruskal, "A nonparametric test for the several sample problem," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 525–540.

[8] E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.

[9] E. L. Lehmann, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24, pp. 23–43.

[10] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 217–226.

[11] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), pp 50–60.

[12] A. Mood, *Introduction to the Theory of Statistics*, McGraw-Hill, 1950.

[13] A. Mood, "On the asymptotic efficiency of certain nonparametric two-sample tests," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 514–522.

[14] G. E. Noether, "Asymptotic properties of the Wald-Wolfowitz test of randomness," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 231–246.

[15] P. B. Patnaik, "The noncentral χ^2 and F distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.

[16] E. J. G. Pitman, Unpublished lecture notes, Columbia University, 1948.

[17] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, 1952.

[18] H. Scheffé, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434–438.

[19] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Statistical Research Memoirs*, Vol. II (1938), University College, London, pp. 126–149.

[20] W. A. Wallis, and W. H. Kruskal, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 583–621.

[21] J. Westenberg, "Significance test for median and interquartile range in samples from continuous populations of any form," *Nederl. Akad. Wetensch., Proc.*, Vol. 51 (1948), pp. 253–261.

[22] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, Vol. 1 (1945), pp. 80–83.

MULTIDIMENSIONAL STOCHASTIC APPROXIMATION METHODS[1]

By Julius R. Blum

University of California, Berkeley and Indiana University

1. Summary. Multidimensional stochastic approximation schemes are presented, and conditions are given for these schemes to converge a.s. (almost surely) to the solutions of k stochastic equations in k unknowns and to the point where a regression function in k variables achieves its maximum.

2. Introduction. Let $H(y \mid x)$ be a family of distribution functions depending upon a real parameter x and let $M(x) = \int_{-\infty}^{\infty} y \, dH(y \mid x)$ be the regression function corresponding to the family $H(y \mid x)$. Robbins and Monro [1] define a stochastic approximation method to solve the equation $M(x) = \alpha$, where α is a specified constant. Their method is such that the approximating random variables converge in probability to θ, where θ is a root of the equation $M(x) = \alpha$. These results are generalized by Wolfowitz [2]. Kiefer and Wolfowitz [3] define a stochastic approximation scheme which converges in probability to θ, where θ is the point at which $M(x)$ achieves a maximum. Finally, it is shown [4] that in fact, in both of the situations mentioned above, the approximating sequence of random variables converges a.s. to θ.

The object of this paper is to extend these results to several dimensions. More precisely we consider the following two problems.

(A) Let $\{Y^{(1)}_{x_1,\cdots,x_k}\}, \cdots, \{Y^{(k)}_{x_1,\cdots,x_k}\}$ be k families of random variables with corresponding families of distribution functions $\{F^{(1)}_{x_1,\cdots,x_k}\}, \cdots, \{F^{(k)}_{x_1,\cdots,x_k}\}$, each depending on k real variables $(x_1, \cdots, x_k)$. Let $M^{(i)}(x_1, \cdots, x_k) = \int_{-\infty}^{\infty} y \, dF^{(i)}_{x_1,\cdots,x_k}$, for $i = 1, \cdots, k$, be the corresponding regression functions. Then, if $\alpha_1, \cdots, \alpha_k$ are k specified numbers, it is desired to find a stochastic approximation method such that the sequence of approximating random vectors converges a.s. to a solution of the equation

$$M^{(i)}(x_1, \cdots, x_k) = \alpha_i, \qquad i = 1, \cdots, k.$$

Here it is assumed that the distributions $F^{(i)}$ and the regression functions $M^{(i)}$ are unknown; however, it is possible to make an observation on the random variable $Y^{(i)}_{x_1,\cdots,x_k}$ for $i = 1, \cdots, k$, and any choice of real numbers $(x_1, \cdots, x_k)$.

(B) Let $\{Y_{x_1,\cdots,x_k}\}$ be a family of random variables, $F_{x_1,\cdots,x_k}$ be the corresponding distribution functions, and $M(x_1, \cdots, x_k)$ the corresponding regres-

Received 10/12/53, revised [illegible]4.

[1] This paper was prepared with the partial support of the Office of Ordnance Research, U.S. Army.

sion function. Subject to the assumptions of (A), it is desired to estimate that set of numbers $(\theta_1, \cdots, \theta_k)$ for which the function M achieves its maximum.

The approximating sequences defined in this paper are straightforward generalizations of the sequences defined in [1] and [3]. The methods of proof used here were strongly motivated by the methods used in [2] and [3].

3. A theorem on almost sure convergence. The following theorem is an immediate consequence of the martingale convergence theorem of Doob [5].

THEOREM. *Let X_n be a sequence of random variables satisfying*

(i) $\sup_n E\{|X_n|\} < \infty$,
(ii) $\sum_{n=1}^{\infty} E\{[E\{X_{n+1} - X_n \mid X_1, \cdots, X_n\}^+]\} < \infty$.

Then X_n converges a.s. *to a random variable.*

As usual, we define X^+ by $X^+ = \frac{1}{2}[X + |X|]$. We immediately obtain the following

COROLLARY. *Let X_n be a sequence of integrable random variables which satisfy condition* (ii) *of the theorem and are bounded below uniformly in n. Then X_n converges* a.s. *to a random variable.*

PROOF. Let $Y_n = X_n - a$, where a is chosen so that $Y_n \geqq 0$ for all n. Then

$$E\{|Y_n|\} = E\{Y_n\} = E\{Y_1\} + \sum_{j=1}^{n-1} E\{Y_{j+1} - Y_j\} \leqq E\{Y_1\} + \sum_{j=1}^{n-1} E\{[E\{X_{j+1} - X_j \mid X_1, \cdots, X_j\}^+]\}.$$

Hence the theorem applies to the sequence Y_n and consequently to the sequence X_n.

4. Convergent sequences of random vectors. Let E_k be a real k-dimensional vector space spanned by the orthogonal unit vectors $u_1, \cdots, u_k$. If x and y are two vectors in E_k, we denote their inner product by $\langle x, y\rangle$ and their norms by $\|x\|$ and $\|y\|$, respectively. Suppose that to each $x \in E_k$ corresponds a random vector $Y_x \in E_k$. Denote by $M(x)$ the vector representing the conditional expectation of Y_x when x is fixed.

Let now $f(x)$ be a real-valued function defined on E_k and possessing continuous partial derivatives of the first and second order. The vector of first partial derivatives will be denoted by $D(x)$ and the matrix of second partial derivatives by $A(x)$. That is

$$D(x) = \left(\frac{\partial f}{\partial x_i}\right) \Big| \, x, \qquad A(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right) \Big| \, x.$$

Then, for any real number a, we have by Taylor's theorem

$$f(x + aY_x) = f(x) + a\langle D(x), Y_x\rangle + \tfrac{1}{2}a^2\langle Y_x, A(x + \theta a Y_x)Y_x\rangle,$$

where θ is a real number with $0 \leqq \theta \leqq 1$. Consequently we may take expectations on both sides to obtain

$$E\{f(x + aY_x)\} = f(x) + a\langle D(x), M(x)\rangle + \tfrac{1}{2}a^2E\{\langle Y_x, A(x + \theta aY_x)Y_x\rangle\}. \tag{4.1}$$

Let now $\{a_n\}$ be a sequence of positive numbers and consider the following sequence of recursively defined random vectors

$$X_{n+1} = X_n + a_nY_n, \tag{4.2}$$

where X_1 is chosen arbitrarily and where Y_n has the distribution of Y_x when X_n yields the observation x. The object of this section is to set down conditions under which X_n converges a.s. to zero.

To simplify writing we shall employ the following notation throughout:

$$Z_x = f(x), \qquad U(x) = \langle D(x), M(x)\rangle, \qquad V_a(x) = E\{\langle Y_x, A(x + \theta aY_x)Y_x\rangle\}.$$

When we substitute the random variables X_n for x and the numbers a_n for a, the corresponding random variables will be denoted by Z_n, U_n, and V_n. We shall assume throughout that $M(0) = 0$.

Consider now the following set A of conditions:

A₁: $$\sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty;$$

A₂: $$Z_x \geqq 0;$$

A₃: $$\sup_{\epsilon \leqq \|x\|} U(x) < 0 \qquad \text{for every } \epsilon > 0;$$

A₄: $$\inf_{\epsilon \leqq \|x\|} |Z_x - Z_0| > 0 \qquad \text{for every } \epsilon > 0;$$

A₅: $$V_a(x) \leqq V < \infty \qquad \text{for every number } a.$$

Then we have

THEOREM 1. *If the sequence a_n satisfies* A₁ *and if there exists a real-valued function $f(x)$ with continuous first and second partial derivatives satisfying* A₂ ⋯ A₅, *then the sequence $\{x_n\}$ defined by* (4.2) *converges a.s. to zero.*

PROOF. From (4.1) we obtain

$$E\{Z_{n+1} \mid Z_1, \cdots, Z_n\} = Z_n + a_nE\{U_n \mid Z_1, \cdots, Z_n\} + \frac{a_n^2}{2}E\{V_n \mid Z_1, \cdots, Z_n\} \text{ a.s.} \tag{4.3}$$

Since $M(0) = 0$, we have, by virtue of conditions A,

$$E\{U_n \mid Z_1, \cdots, Z_n\} \leqq 0 \text{ a.s.}, \qquad E\{V_n \mid Z_1, \cdots, Z_n\} \leqq V \text{ a.s.},$$

both for all n. Hence

$$E\{Z_{n+1} - Z_n \mid Z_1, \cdots, Z_n\} \leqq \tfrac{1}{2}a_n^2 V \text{ a.s.} \tag{4.4}$$

We may assume V to be nonnegative. Using this fact together with conditions A_1 and A_2, we may apply the corollary of Section 3 to obtain

$$P\{Z_n \text{ converges}\} = 1. \tag{4.5}$$

Taking expectations on both sides of (4.3) and iterating, we have

$$E\{Z_{n+1}\} = Z_1 + \sum_{j=1}^{n} a_j E\{U_j\} + \sum_{j=1}^{n} \tfrac{1}{2} a_j^2 E\{V_j\}.$$

From what has been said above it follows that

$$E\{Z_n\} \geqq 0, \qquad E\{U_n\} \leqq 0, \qquad E\{V_n\} \leqq V, \qquad n = 1, \cdots.$$

Since V is nonnegative and the series $\sum_1^\infty a_n^2$ converges, the nonpositive term series $\sum_1^\infty a_n E\{U_n\}$ also converges. By virtue of the fact that $\sum_1^\infty a_n = \infty$ we have

$$\limsup_{n\to\infty} E\{U_n\} = 0, \qquad \liminf_{n\to\infty} E\{| U_n |\} = 0.$$

Let $\{n_k\}$ be an infinite sequence of integers such that $\lim_{k\to\infty} E\{|U_{n_k}|\} = 0$. Then U_{n_k} converges to zero in probability and there exists a further subsequence say $\{U_{m_k}\}$ such that

$$P\{\lim_{k\to\infty} U_{m_k} = 0\} = 1.$$

From condition A_3 it follows that $P\{\lim_{k\to\infty} X_{m_k} = 0\} = 1$. Since Z_n is a continuous function of X_n it follows from (4.5) that

$$P\{\lim_{n\to\infty} Z_n = Z_0\} = 1. \tag{4.6}$$

Now consider a sample sequence $\{X_n\}$ such that for the corresponding sequence $\{Z_n\}$ we have $\lim_{n\to\infty} Z_n = Z_0$. From condition A_4 it is clear that for such a sequence we must have $\lim_{n\to\infty} X_n = 0$. Hence (4.6) gives the desired result.

We may obtain the same result by assuming a slightly different set of conditions: A′, changing A_3 and A_5 to:

A_3': There exists $\epsilon > 0$ such that $\sup_{0 \leqq ||x|| < \epsilon} V_a(x) \leqq V < \infty$ for every number a;

A_5': There exists $\lambda > 0$, with $\lambda > \frac{1}{2}a_n$ for each n, such that $\sup_{\delta \leqq ||x||} [U(x) + \lambda V_a^+(x)] < 0$ for every $\delta > 0$ and every number a.

Then we have

THEOREM 2. *If the sequence $\{a_n\}$ satisfies condition A_1 and if there exists a real-valued function $f(x)$ with continuous first and second partial derivatives satisfying A_2, A_3', A_4, and A_5, then the sequence $\{X_n\}$ defined by* (4.2) *converges* a.s. *to zero.*

The proof of this theorem follows very closely that of Theorem 1, and so is omitted.

5. Examples. In this section we illustrate the results of the previous section by a few simple examples. Assume that the problem is as described in (A) of Section 2. Then to each $x \varepsilon E_k$ corresponds to a random vector $Y_x \varepsilon E_k$ with coordinates $Y_x^{(i)}$ for $i = 1, \cdots, k$. Let $M(x)$ be the vector of conditional expectations, when x is given. Without loss of generality we assume that $\alpha_i = \theta_i = 0$ for $i = 1, \cdots, k$.

Example I. Let B be a negative definite $k \times k$ matrix and assume

(i) for some $\rho > 0$, $\|x\| \leqq \rho$ implies $M(x) = Bx$;
(ii) $\|x\| > \rho$ implies $M(x) = M([\rho/\|x\|]x)$;
(iii) $\sigma_x^{2(i)} \leqq \sigma^2 < \infty$ for each $x \varepsilon E_k$, and each $i = 1, \cdots, k$, where $\sigma_x^{2(i)}$ is the variance of the ith component of Y_x.

Under these conditions it is clear that both $\|M(x)\|$ and $E\{\|Y_x\|^2\}$ are bounded uniformly in x. Now define $f(x)$ by $f(x) = \|x\|^2$. If we choose the sequence $\{a_n\}$ to satisfy condition A_1, we can easily verify that the remainder of condition A is satisfied. To do this we note that A_2 and A_4 are obviously satisfied from the choice of $f(x)$. Further we have

$$U(x) = \begin{cases} 2\langle x, Bx\rangle & \|x\| \leqq \rho; \\ 2[\rho/\|x\|]\,\langle x, Bx\rangle & \|x\| > \rho; \end{cases}$$

$$V_a(x) = 2E\{\|Y_x\|^2\} \qquad \text{for every number } a.$$

From the boundedness of $E\{\|Y_x\|^2\}$ it is clear that A_5 if also satisfied. It remains to check A_3. To do this we recall that for every negative definite matrix B there exists a positive number b such that $\langle x, Bx\rangle \leqq -b\|x\|^2$. Thus if ϵ is any positive number with $0 < \epsilon \leqq \rho$, we have

$$\langle x, Bx\rangle \leqq -b\epsilon^2 \text{ if } \epsilon \leqq \|x\| \leqq \rho, \qquad [\rho/\|x\|]\,\langle x, Bx\rangle \leqq -b\rho^2 \text{ if } \|x\| > \rho.$$

Hence A_3 is also satisfied and Theorem 1 applies.

Example II. Consider a negative definite matrix B and assume

(i) $M(x) = Bx$;
(ii) there exist $\epsilon > 0$ and $C > 0$ such that $\|x\| \leqq \epsilon$ implies $E\{\|Y_x\|^2\} \leqq C$
(iii) there exists $\rho > 0$ such that $\|x\| > \epsilon$ implies

$$\langle x, Bx\rangle + \rho E\{\|Y_x\|^2\} \leqq 0.$$

With $f(x)$ again defined by $f(x) = \|x\|^2$, we have

$$U(x) = 2\langle x, Bx\rangle, \qquad V_a(x) = 2E\{\|Y_x\|^2\} \qquad \text{for all } a.$$

Hence it is clear that if we choose the sequence $\{a_n\}$ to satisfy condition A_1, we need only verify A_5', since the other conditions follow immediately. To do this, assume first that $\|x\| \leqq \epsilon$ as determined by assumption (ii) of this example, and let λ be any positive number. Let $b > 0$ be such that $\langle x, Bx\rangle \leqq -b\|x\|^2$.

Then we have

$$U(x) + \lambda V^+(x) = 2[\langle x, Bx\rangle + E\{\| Y_x \|^2\}]$$
$$\leqq 2[\langle x, Bx\rangle + \lambda C] \leqq 2[-b \| x \|^2 + \lambda C].$$

Hence it is clear that if $0 < \delta \leqq \| x \| \leqq \epsilon$, we can choose λ_1 such that

$$U(x) + \lambda_1 V^+(x) \leqq 2[-b\delta^2 + \lambda_1 C] < 0,$$

and if $\| x \| > \epsilon$, choose $0 < \lambda_2 < \rho$, where ρ is determined by assumption (iii) of the example. Then

$$\frac{[U(x) + \lambda_2 V^+(x)]}{2} = \left(\frac{\rho - \lambda_2}{\rho}\right)\langle x, Bx\rangle + \frac{\lambda_2}{\rho}[\langle x, Bx\rangle + \rho E\{\| Y_x\|^2\}]$$
$$\leqq \left(\frac{\rho - \lambda_2}{\rho}\right)\langle x, Bx\rangle \leqq -\left(\frac{\rho - \lambda_2}{\rho}\right) b\epsilon^2 < 0.$$

Hence by choosing $\lambda = \min(\lambda_1, \lambda_2)$ we satisfy condition A_4 and Theorem 2 applies.

6. The maximum of a regression function in several variables. In this section we turn to problem (B) of Section 2. Assume once more that x is a variable point in E_k and to each x corresponds a random variable Y_x, with corresponding regression function $M(x)$. Assume, without loss of generality, that $M(x)$ has a unique maximum at $x = 0$. The problem becomes one of constructing a sequence $\{X_n\}$ of random vectors with the property

$$P\{\lim_{n\to\infty} X_n = 0\} = 1.$$

Let $\{a_n\}$ and $\{c_n\}$ be two infinite sequences of positive numbers satisfying conditions B:

$$B_1: \lim_{n\to\infty} c_n = 0, \qquad B_2: \sum_{n=1}^{\infty} a_n = \infty, \qquad B_3: \sum_{n=1}^{\infty} a_n c_n < \infty,$$
$$B_4: \sum_{n=1}^{\infty} \left(\frac{a_n}{c_n}\right)^2 < \infty,$$

Suppose now $x \varepsilon E_k$ and let c be a positive number. Let $u_1, \cdots, u_k$ be the orthonormal set spanning E_k. We construct a random vector $Y_{x,c}$ by taking $k + 1$ independent observations on the random variables $Y_x, Y_{x+cu_1}, \cdots, Y_{x+cu_k}$ and defining

$$Y_{x,c} = [(Y_{x+cu_1} - Y_x), \cdots, (Y_{x+cu_k} - Y_x)].$$

We proceed to construct a recursive sequence of random vectors by choosing X_1 arbitrarily and defining

$$X_{n+1} = X_n + a_n Y_n/c_n, \tag{6.1}$$

where Y_n has the distribution of Y_{x,c_n} when X_n yields the observation x. The intuitive reason for (6.1) is fairly clear, since Y_n/c_n is the vector in the direction

of the maximum slope of the plane determined by the $k + 1$ vectors

$$(X_n, \ Y_{X_n}), \ (X_n + c_n u_1, \ Y_{X_n + c_n u_1}), \ \cdots, \ (X_n + c_n u_k, \ Y_{X_n + c_n u_k}).$$

We denote the vector of first partial derivatives and the matrix of second partial derivatives of $M(x)$ by $D(x)$ and $A(x)$, respectively. We write D_n for $D(X_n)$ and A_n for $A(X_n)$, and denote by $\bar{A}_n$ the vector whose coordinates are the diagonal entries of A_n, by Δ_n the vector $E\{Y_n \mid X_n\}$, and by σ_x^2 the variance of Y_x. Without loss of generality we assume that $M(0) = 0$ so that $M(x) \leqq 0$ for all x. Then we have

THEOREM 3. *Suppose the sequences $\{a_n\}$ and $\{c_n\}$ satisfy conditions B and further that*

(i) *$M(x)$ is continuous with continuous first and second derivatives;*

(ii) $\sigma_x^2 \leqq \sigma^2 < \infty$;

(iii) *for every positive number ϵ there exists a positive number $\rho(\epsilon)$ such that*

$$\| x \| \geqq \epsilon \quad \text{implies} \quad M(x) \leqq -\rho(\epsilon) \quad \text{and} \quad \| D(x) \| \geqq \rho(\epsilon).$$

(iv) *The second partial derivatives $\partial^2 M(x)/\partial x_i \partial x_j$ are bounded for $i, j = 1, \cdots, k$.*

Then the sequence $\{X_n\}$ defined by (6.1) converges a.s. to zero.

PROOF. Expanding $-M(X_{n+1})$ we obtain, with $0 \leqq \theta \leqq 1$,

$$-M(X_{n+1}) = -M(X_n) - \frac{a_n}{c_n} \langle D_n, Y_n \rangle - \frac{a_n^2}{2c_n^2} \langle Y_n, A\left(X_n + \theta \frac{a_n}{c_n} Y_n\right) Y_n \rangle.$$

Taking conditional expectation for given X_n we have

$$E\{-M(X_{n+1}) \mid X_n\} = -M(X_n) - \frac{a_n}{c_n} \langle D_n, \Delta_n \rangle - \frac{a_n^2}{2c_n^2}$$

$$E\left\{\langle Y_n, A\left(X_n + \theta \frac{a_n}{c_n} Y_n\right) Y_n \rangle \mid X_n\right\} \text{ a.s.}$$

Since $A(x)$ is a bounded matrix and σ_x^2 is bounded, we have

$$\left| E\left\{\langle Y_n, A\left(X_n + \theta \frac{a_n}{c_n} Y_n\right) Y_n \rangle \mid X_n\right\} \right| \leqq K_1 \| \Delta_n \|^2 + K_2,$$

where K_1 and K_2 are suitably chosen positive constants. By virtue of the hypothesis we obtain

$$\Delta_n^{(i)} = c_n \langle D_n, u_i \rangle + \tfrac{1}{2} c_n^2 \langle u_i, A(X_n + \theta^{(i)} c_n u_i) u_i \rangle, \quad i = 1, \cdots, k,$$

where $\Delta_n^{(i)}$ is the ith component of Δ_n and $0 \leqq \theta^{(i)} \leqq 1$ for $i = 1, \cdots, k$. Hence

$$\langle D_n, \Delta_n \rangle = c_n \| D_n \|^2 + \tfrac{1}{2} c_n^2 \langle D_n, \bar{A}_n \rangle$$

$$\| \Delta_n \|^2 = c_n^2 \| D_n \|^2 + c_n^3 \langle D_n, \bar{A}_n \rangle + \tfrac{1}{4} c_n^4 \| \bar{A}_n \|^2.$$

Now by hypothesis, $\| \bar{A}_n \|$ is bounded, say $\| \bar{A}_n \|^2 \leqq K_3$. Then

$$\| \langle D_n, A_n \rangle \|^2 \leqq K_3 \| D_n \|^2.$$

After some computation we find

$$E\{-M(X_{n+1}) \mid X_n\} \leqq -M(X_n) - a_n\{\|D_n\|^2[1 - \tfrac{1}{2}K_1a_n] - \|D_n\| K_3^{1/2}[\tfrac{1}{2}c_n - \tfrac{1}{2}K_1a_nc_n]\} + \tfrac{1}{8}K_1K_3a_n^2c_n^2 + \tfrac{1}{2}K_2a_n^2/c_n^2 \quad \text{a.s.},$$

where n is chosen so large that $[1 - \frac{1}{2}K_1a_n]$ and $[c_n - K_1a_nc_n]$ are both nonnegative.

Let λ_n be a sequence of random variables defined by

$$\lambda_n = \begin{cases} 1 & \text{if } \|D_n\| \geqq 1, \\ 0 & \text{otherwise.} \end{cases}$$

We note that for n sufficiently large we have

$$a_n\{\|D_n\|^2[1 - \tfrac{1}{2}K_1a_n] - \lambda_n \|D_n\| K_3^{1/2}[\tfrac{1}{2}c_n - \tfrac{1}{2}K_1a_nc_n]\} \geqq 0. \tag{6.2}$$

Hence, for such n we obtain

$$E\{-M(X_{n+1}) \mid X_n\} \leqq -M(X_n) + a_nc_n(1 - \lambda_n) \|D_n\| K_3^{1/2} \mid \tfrac{1}{2} - \tfrac{1}{2}K_1a_n \mid + \tfrac{1}{8}K_1K_3a_n^2c_n^2 + \tfrac{1}{2}K_2a_n^2/c_n^2 \quad \text{a.s.}$$

This inequality clearly is still preserved if we take conditional expectations with respect to $M(X_n)$ on both sides. But now we note that

$\sum_{j=1}^n a_jc_jK_3^{1/2}[\frac{1}{2} - \frac{1}{2}K_1a_n] \, E\{(1 - \lambda_n) \|D_n\| \mid M(X_n)\}$ converges a.s.;

$\sum_1^n \frac{1}{8}K_1K_3a_j^2c_j^2$ and $\sum_1^n \frac{1}{2}K_2a_j^2/c_j^2$ both converge.

These follow from conditions B and the definitions of λ_n. Hence, we may again apply the corollary of Section 3 to obtain that $M(X_n)$ converges a.s. to a random variable. Now we note that $\sum_1^n a_j$ diverges to $+\infty$ and that $M(X_n) \leqq 0$. Hence the series

$$\sum_{j=1}^n a_j E\{\|D_j\|^2 [1 - \tfrac{1}{2}K_1a_j] - \lambda_j \|D_j\| K_3^{1/2}[\tfrac{1}{2}c_j - \tfrac{1}{2}K_1a_jc_j\}$$

converges. This, together with (6.2), insures the existence of a subsequence D_{n_k} with the property $P\{\lim_{k\to\infty} D_{n_k} = 0\} = 1$. Hence X_{n_k} converges a.s. to zero. Since $M(x)$ is continuous and $M(0) = 0$, we have $P\{\lim_{k\to\infty} M(X_n) = 0\} = 1$, which implies the desired result.

REFERENCES

[1] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400–407.

[2] J. Wolfowitz, "On the stochastic approximation method of Robbins and Monro," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 457–461.

[3] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462–466.

[4] J. R. Blum, "Approximation methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382–386.

[5] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, (1953).

CERTAIN INEQUALITIES IN INFORMATION THEORY AND THE CRAMÉR-RAO INEQUALITY

By S. Kullback

The George Washington University

1. Summary and Introduction. The Cramér-Rao inequality provides, under certain regularity conditions, a lower bound for the variance of an estimator [7], [15]. Various generalizations, extensions and improvements in the bound have been made, by Barankin [1], [2], Bhattacharyya [3], Chapman and Robbins [5], Fraser and Guttman [11], Kiefer [12], and Wolfowitz [16], among others.

Further considerations of certain inequality properties of a measure of information, discussed by Kullback and Leibler [14], yields a greater lower bound for the information measure (formula (4.11)), and leads to a result which may be considered a generalization of the Cramér-Rao inequality, the latter following as a special case. The results are used to define discrimination efficiency and estimation efficiency at a point in parameter space.

2. The first inequality. We use the notation and terminology of [14]. Consider the measurable transformations T_N of the probability spaces $(\mathfrak{X}, \mathcal{S}, \mu_i)$ onto the probability spaces $(\mathcal{Y}, \mathfrak{J}, \nu_i^{(N)})$, and suppose for $G \varepsilon \mathfrak{J}$ that $\nu_i^{(N)}(G) = \mu_i(T_N^{-1} G)$ for $i = 1$ or 2.

Theorem 2.1. *Let the T_N be such that*

$$\lim_{N\to\infty} \nu_i^{(N)}(G) = \nu_i(G), \qquad i = 1, 2; \qquad G \varepsilon \mathfrak{J}, \tag{2.1}$$

Then

$$I(1:2; x) \geqq \liminf_{N\to\infty} I'_N(1:2; y) \geqq I'(1:2; y); \tag{2.2}$$

$$J(1, 2; x) \geqq \liminf_{N\to\infty} J'_N(1, 2; y) \geqq J'(1, 2; y). \tag{2.2'}$$

Proof. We first derive a result which is similar to a lemma used by Doob [8]. Using Lemma 3.2 of [14], we have

$$I'_N(1:2; y) \geqq \sum \nu_1^{(N)}(G_j) \log \frac{\nu_1^{(N)}(G_j)}{\nu_2^{(N)}(G_j)}, \tag{2.3}$$

where the sum is taken over any set of pairwise disjunct G_j such that $\bigcup_j G_j = \mathcal{Y}$. Accordingly,

$$\liminf_{N\to\infty} I'_N(1:2; y) \geqq \sum \nu_1(G_j) \log \frac{\nu_1(G_j)}{\nu_2(G_j)}, \tag{2.4}$$

Received 10/12/53, revised 4/20/54.

and therefore

$$\liminf_{N\to\infty} I'_N(1:2; y) \geqq I'(1:2; y), \tag{2.5}$$

since the right member of (2.5) is the l.u.b. of the right member of (2.4). In conjunction with Theorem 4.1 and paragraph 5 of [14] and (2.5), the inequalities (2.2) and (2.2′) follow. These are used herein only in Section 3.

3. An example. Consider N independent observations from the binomial distributions $B(p_i, q_i)$, for $i = 1$, or 2, which as $N \to \infty$ approach as limits the Poisson exponential distributions with means $m_i = Np_i$, for $i = 1$ or 2. It may be verified readily that

$$\begin{aligned} I'_N(1:2; y) &= \sum \frac{N!}{y!(N-y)!} p_1^y q_1^{N-y} \log \frac{p_1^y q_1^{N-y}}{p_2^y q_2^{N-y}} \\ &= N\left(p_1 \log \frac{p_1}{p_2} + q_1 \log \frac{q_1}{q_2}\right), \end{aligned} \tag{3.1}$$

$$I'(1:2; y) = \sum \frac{m_1^y e^{-m_1}}{y!} \log \frac{m_1^y e^{-m_1}}{m_2^y e^{-m_2}} = (m_2 - m_1) + m_1 \log \frac{m_1}{m_2}. \tag{3.2}$$

Using the well known inequality $x_1 \log (x_1/x_2) \geqq x_1 - x_2$, and $m_i = Np_i$ for $i = 1$ or 2, it is found that

$$\begin{aligned} Np_1 \log \frac{p_1}{p_2} + Nq_1 \log \frac{q_1}{q_2} &= m_1 \log \frac{m_1}{m_2} + N\left(1 - \frac{m_1}{N}\right) \log \frac{1 - m_1/N}{1 - m_2/N} \\ &\geqq m_1 \log \frac{m_1}{m_2} + N\left(\frac{m_2}{N} - \frac{m_1}{N}\right) = m_1 \log \frac{m_1}{m_2} + (m_2 - m_1), \end{aligned} \tag{3.3}$$

or

$$\liminf_{N\to\infty} I'_N(1:2; y) \geqq I'(1:2; y). \tag{3.4}$$

As a matter of fact, for this particular case, as may be readily seen from the first two members of (3.3),

$$\lim_{N\to\infty} I'_N(1:2; y) = I'(1:2; y). \tag{3.5}$$

4. The second inequality. Suppose $g_1(y)$, $g_2(y)$, and $g^*(y)$ are densities satisfying the conditions of paragraph 4 of [14]. Then using Lemma 3.1 of [14],

$$\begin{aligned} \int g_1(y) \log \frac{g_1(y)}{g_2(y)}\, d\gamma(y) &+ \int g_1(y) \log \frac{g_2(y)}{g^*(y)}\, d\gamma(y) \\ &= \int g_1(y) \log \frac{g_1(y)}{g^*(y)}\, d\gamma(y) \geqq 0, \end{aligned} \tag{4.1}$$

or

$$\int g_1(y) \log \frac{g_1(y)}{g_2(y)}\, d\gamma(y) \geqq \int g_1(y) \log \frac{g^*(y)}{g_2(y)}\, d\gamma(y). \tag{4.2}$$

In particular, let us take, for real t,

$$g^*(y) = \frac{e^{ty} g_2(y)}{M_2(t)}, \qquad M_2(t) = \int e^{ty} g_2(y)\, d\gamma(y), \tag{4.3}$$

so that (4.2) becomes

$$I'\,(1:2;\, y) \geqq at - \log M_2(t), \qquad a = E_1(y), \tag{4.4}$$

with equality if and only if

$$g_1(y) = g^*(y) = \frac{e^{ty} g_2(y)}{M_2(t)} \quad [\gamma(y)]. \tag{4.5}$$

To investigate further the right member of (4.4) we will use the notation, and, in particular, the results of paragraphs 4 and 6 of Chernoff [6]. Clearly

$$I'(1:2;\, y) \geqq \sup_t\, (at - \log M_2(t)) = -\log m_2(a), \tag{4.6}$$

where $m_2(a) = \inf_t e^{-at} M_2(t)$. Note that for the value of t satisfying $a = N_2(t(a)) / M_2(t(a))$, we have

$$-\log m_2(a) = \int g^*(y) \log \frac{g^*(y)}{g_2(y)}\, d\gamma(y) \geqq 0.$$

From this, or the results of Lemma 7 of Chernoff [6], it follows that $-\log m_2(a)$ is a convex function of a. Limiting ourselves to statistics y for which $E_2(y)$ and $\mathrm{Var}_2(y)$ are finite, the results of Chernoff [6] may also be derived for the case $a \geqq E_2(y)$.

We can write

$$\log m_2(a) = \log m_2(E_2(y)) + (a - E_2(y)) \frac{d}{da} \log m_2(a) \Big|_{a=E_2(y)} + \frac{(a - E_2(y))^2}{2!} \left(-\frac{dt(b)}{db} \right), \tag{4.7}$$

where b is between a and $E_2(y)$. But as Chernoff [6] has shown,

$$\log m_2\,(E_2(y)) = 0, \qquad \frac{d}{da} \log m_2(a) \Big|_{a=E_2(y)} = 0,$$
$$\frac{dt(a)}{da} \Big|_{a=E_2(y)} = \frac{1}{\mathrm{Var}_2(y)}, \qquad \frac{dt(b)}{db} = \frac{1}{\mathrm{Var}_*(y)}, \tag{4.8}$$

where $\mathrm{Var}_*(y)$ is the variance of y for the distribution defined by

$$g_*(y) = e^{t(b)y} g_2(y) \,/\, M_2(t(b)). \tag{4.9}$$

From (4.6), (4.7), and (4.8) it follows that

$$I'(1:2;\, y) \geqq (E_1(y) - E_2(y))^2 \,/\, 2\, \mathrm{Var}_*(y), \tag{4.10}$$

where [13] the right side is the value of $I(1:2)$ for two normal distributions with common variance $\mathrm{Var}_*(y)$ and means $E_1(y)$ and $E_2(y)$.

We take y as the linear function $y = c_1y_1 + c_2y_2 + \cdots + c_ky_k$, where the random variables $y_1, y_2, \cdots, y_k$ are such that the requirements already imposed on y are satisfied and $\text{Var}_*(y) = \sum_{i,j=1}^k c_ic_j \text{ cov}_*(y_i, y_j)$. Then, as is known [13], the l.u.b. of the right member of (4.10) for possible values of the c's is given by the quadratic form $\frac{1}{2}\delta'\sigma_*^{-1}\delta$, where δ is the one column matrix of the differences $\delta_j = E_1(y_j) - E_2(y_j)$ for $j = 1, 2, \cdots, k$, and δ' is the transpose of δ while σ_* is the matrix of variances and covariance of the y_j for $j = 1, 2, \cdots, k$ in the distribution defined by (4.9).

We thus have the second inequality

$$I(1:2; x) \geqq I'(1:2; y) \geqq \tfrac{1}{2}\delta'\sigma_*^{-1}\delta. \tag{4.11}$$

For the binomial distribution, this yields

$$p_1 \log \frac{p_1}{p_2} + q_1 \log \frac{q_1}{q_2} \geqq \frac{(p_1 - p_2)^2}{2p_* q_*}, \quad p_* = \frac{p_2e^t}{p_2e^t + q_2}, \quad q_* = \frac{q_2}{p_2e^t + q_2}, \tag{4.12}$$

for some value of t between 0 (when $p_* = p_2$) and $\log p_1q_2/q_1p_2$ (when $p_* = p_1$). Note that $p_* = b$, and that from our derivation b is between p_1 and p_2.

5. The Cramér-Rao inequality. For the parametric case, where the populations are neighboring points θ and $\theta + \Delta\theta$ in the k-dimensional parameter space and the y_j for $j = 1, 2, \cdots, k$ are unbiased estimators of the parameters, (4.11) yields, under suitable regularity conditions [14],

$$(\Delta\theta)'G(\Delta\theta) \geqq (\Delta\theta)'H(\Delta\theta) \geqq (\Delta\theta)'\sigma^{-1}(\Delta\theta), \tag{5.1}$$

where $\Delta\theta$ is the one column matrix of the $\Delta\theta_j$ for $j = 1, 2, \cdots, k$ and $(\Delta\theta)'$ is its transpose, while G and H are respectively the matrices $(g_{\alpha\beta})$ and $(h_{\alpha\beta})$, for $\alpha, \beta = 1, 2, \cdots, k$, where

$$g_{\alpha\beta} = \int f(x) \left(\frac{\partial}{\partial\theta_\alpha} \log f(x)\right)\left(\frac{\partial}{\partial\theta_\beta} \log f(x)\right) d\lambda(x),$$

$$h_{\alpha\beta} = \int g(y) \left(\frac{\partial}{\partial\theta_\alpha} \log g(y)\right)\left(\frac{\partial}{\partial\theta_\beta} \log g(y)\right) d\gamma(y),$$

and σ is the matrix of variances and covariances of the estimators.

It should be observed that the discussion in Sections 4 and 5 holds whether we are dealing with a fixed sample size or sequential procedure. For the latter case, ([16] p. 216) let $\mathfrak{X}$ of the probability spaces $(\mathfrak{X}, \mathcal{S}, \mu_i)$, be the space of all possible infinite sequences (x) of observations $x_1, x_2, \cdots$. Let there be given an infinite sequence of Borel measurable functions $\phi_1(x_1), \phi_2(x_1, x_2), \cdots, \phi_j(x_1, x_2, \cdots, x_j)$ $\cdots$, defined for all observable sequences in $\mathfrak{X}$ such that each takes only the values zero and one. We further assume that at least one of the functions $\phi_1(x_1)$, $\phi_2(x_1, x_2), \cdots$ takes the value one $[\lambda(x)]$, and let n be the smallest integer for which this occurs. Thus $n(x)$ is a chance variable.

The sequential process is then defined as follows. Take an observation and find $\phi_1(x_1)$. If it is unity, the sampling process stops; otherwise sampling con-

tinues. If a second observation is taken and the value of $\phi_2(x_1, x_2)$ is unity, the process stops; otherwise it continues, and so on. In general, after taking j observations,

$$\phi_i(x_1, x_2, \cdots, x_i) = 0 \text{ for } i = 1, 2, \cdots, j - 1. \text{ If}$$

$$\phi_j(x_1, x_2, \cdots, x_j) = 1, \text{ sampling stops; otherwise it is continued.}$$

If R_j denotes the set of all points $(x_1, x_2, \cdots)$ for which the process stops with the jth observation, then $\mathfrak{X} = \bigcup_j R_j$. The variable y is taken as a function of the observations $x_1, x_2, \cdots, x_n$ (those obtained prior to the termination of the process of drawing observations).

Thus the results in (4.11) and (5.1) hold for fixed sample size or sequential procedures.

6. Quadratic forms. Certain useful results with respect to quadratic forms, which are essentially corollaries of known theorems, are needed for the subsequent discussion.

LEMMA 6.1. *If both $X'AX$ and $X'CX$ are positive definite quadratic forms (matric notation) such that $X'AX \geqq X'CX$, then*

(a) *the roots of $|A - \lambda C| = 0$ are real and $\geqq 1$;*

(b) $|A| \geqq |C|$;

(c) *any principal minor of A is not less than the corresponding principal minor of C, (determinant or quadratic form);*

(d) $Y'C^{-1}Y \geqq Y'A^{-1}Y$;

(e) *any principal minor of C^{-1} is not less than the corresponding principal minor of A^{-1} (determinant or quadratic form).*

PROOF. Results (a), (b), and (c) are immediate corollaries of theorems 44 and 48 in Ferrar [10]. Since $A^{-1} = C^{-1}CA^{-1}$ and $C^{-1} = C^{-1}AA^{-1}$, there exists a non-singular matrix B such that (Bôcher [4], p. 301) $C^{-1} = B'AB$ and $A^{-1} = B'CB$. Thus applying the transformation $X = BY$ gives

$$X'AX = Y'B'ABY = Y'C^{-1}Y, \qquad X'CX = Y'B'CBY = Y'A^{-1}Y,$$

and (d) and (e) then follow.

7. Efficiency. With respect to the estimators y_i of Section 5, the *discrimination efficiency* at a point P in the k-dimensional parameter space (P.S.) is defined by

$$\lambda = \frac{(d\theta)' H(d\theta)}{(d\theta)' G(d\theta)}. \tag{7.1}$$

We take $(d\theta)'G(d\theta)$ as the basis of the metric of (P.S.). The $g_{\alpha\beta}$ for $\alpha, \beta = 1, 2, \cdots, k$, are the components of a covariant tensor of the second order which is called the *fundamental tensor* of the metric (Eisenhart [9]). Since $(d\theta)'H(d\theta) \leqq (d\theta)'G(d\theta)$ and both forms are positive definite, the roots of

$$|H - \lambda G| = 0, \tag{7.2}$$

are real, positive, and all $\leqq 1$. Accordingly there exists a real transformation of the θ's such that at a point P in (P.S.) the forms in (7.1) may be written as

$$(7.3) \qquad \lambda = \frac{\lambda_1\, d\psi_1^2 + \cdots + \lambda_k\, d\psi_k^2}{d\psi_1^2 + \cdots + d\psi_k^2}$$

and $\lambda_1, \lambda_2, \cdots, \lambda_k$, are the roots of (7.2) (Eisenhart [9] p. 108). If we write

$$(7.4) \qquad \cos^2 \alpha_i = \frac{d\psi_i^2}{d\psi_1^2 + \cdots + d\psi_k^2}, \qquad i = 1, 2, \cdots, k,$$

then (7.3) may be written as

$$(7.5) \qquad \lambda = \lambda_1 \cos^2 \alpha_1 + \lambda_2 \cos^2 \alpha_2 + \cdots + \lambda_k \cos^2 \alpha_k .$$

The directions at the point P determined by $\cos \alpha_1 = 1$, $\cos \alpha_2 = 1$, $\cdots$, are known as the *principal directions* determined by the tensor $h_{\alpha\beta}$ (Eisenhart [9], p. 110). Furthermore, at the point P the finite maxima and minima of λ defined by (7.1) are given by the principal directions at the point and are indeed the roots of (7.2). Since $(d\theta)'G(d\theta)$ is positive definite, λ is finite for all directions (Eisenhart [9], par. 33).

As the *estimation efficiency* of the estimators $y_1, y_2, \cdots, y_k$, we take the product of the discrimination efficiencies for the principal directions at the point P, that is,

$$(7.6) \qquad \text{Eff} = \lambda_1\lambda_2 \cdots \lambda_k = |H| \,/\, |G| \leqq 1,$$

which is invariant for all nonsingular transformations of the parameters, with equality holding if and only if the estimators are sufficient [14].

8. Asymptotic efficiency. Suppose we have n independent observations from an l-variate population with k parameters. It is also of interest to consider, instead of (7.1), the *asymptotic discrimination efficiency* at a point P in (P.S.) defined by

$$(8.1) \qquad \lambda = \frac{(d\theta)'\sigma^{-1}(d\theta)}{n(d\theta)'\,G(d\theta)}, \qquad n \text{ large},$$

where the elements of the matrix G are computed for a single observation from the l-variate population. Since $(d\theta)'\sigma^{-1}(d\theta) \leqq n(d\theta)'G(d\theta)$ and both forms are positive definite, the roots of

$$(8.2) \qquad |\sigma^{-1} - \lambda n G| = 0$$

are real, positive and $\leqq 1$. As in Section 7, the roots of (8.2) are the finite maxima and minima of (8.1) at a point P in (P.S.) and are given by the principal directions determined by the tensor $\sigma^{\alpha\beta}$ at the point.

As the *asymptotic estimation efficiency* of the unbiased estimators $y_1, y_2, \cdots, y_k$

(cf. Cramér [7], pp. 489, 494) we take the product of the asymptotic discrimination efficiencies for the principal directions at the point P, that is,

$$\text{Asymp Eff} = \lambda_1\lambda_2 \cdots \lambda_k = |\sigma^{-1}| \,/\, |nG| \leqq 1, \qquad n \text{ large},$$

the equality holding for all n if the estimators are sufficient and (4.5) is satisfied. If $|\sigma|\,|G| \to n^{-k}$, then the asymptotic efficiency approaches unity and $\lambda_i \to 1$ for $i = 1, 2, \cdots, k$.

REFERENCES

[1] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 477–501.

[2] E. W. Barankin, "Concerning some inequalities in the theory of statistical estimation," *Skand. Aktuarietids*, Vol. 34 (1951), pp. 35–40.

[3] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1946), pp. 1–14, (1947), pp. 201–218, (1948) pp. 315–328.

[4] M. Bôcher, *Higher Algebra*, MacMillan, 1924.

[5] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 581–586.

[6] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.

[7] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.

[8] J. L. Doob, "Statistical estimation," *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 410–421.

[9] L. P. Eisenhart, *Riemannian Geometry*, Princeton Univ. Press, 1926.

[10] W. L. Ferrar, *Algebra*, Oxford Univ. Press, 1941.

[11] D. A. S. Fraser and I. Guttman, "Bhattacharyya bounds without regularity assumptions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 629–632.

[12] J. Kiefer, "On minimum variance estimators," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 627–629.

[13] S. Kullback, "An application of information theory to multivariate analysis," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 88–102.

[14] S. Kullback and R. A. Leibler, "Information and sufficiency," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 79–86.

[15] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81–91.

[16] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–230.

SOME FURTHER RESULTS IN SIMULTANEOUS CONFIDENCE INTERVAL ESTIMATION

BY S. N. ROY

University of North Carolina

1. Summary. This paper is a follow-up of a previous paper [1], the full implications of some of the results there being brought out here in terms that are physically more meaningful. Two cases of simultaneous confidence bounds, I and II, are given, in each case with a confidence coefficient which is to be greater than or equal to a preassigned level.

Case I relates to the characteristic roots of Σ and $\Sigma_1\Sigma_2^{-1}$, where Σ stands for the dispersion matrix of one p-variate and Σ_1 and Σ_2 for the dispersion matrices of two p-variate normal populations. Case II relates to a $(p + q)$-variate normal population $(p \leqq q)$, for which the matrix of regression of the p-set on the q-set is defined in a natural manner. This matrix is denoted by $\beta(p \times q)$ and simultaneous confidence bounds are given on all bilinear compounds of this matrix (with arbitrary coefficient vectors of unit modulus).

Confidence bounds on the characteristic roots of Σ and $\Sigma_1\Sigma_2^{-1}$ are given respectively by (3.1.3) and (3.2.8). Confidence bounds on the bilinear compounds of the regression matrix β are given by (4.7).

2. Introduction. Let us denote by A' the transpose of a matrix A, and shorten positive definite into p.d. and positive semidefinite into p.s.d. Also let $c_{\min}(M)$ and $c_{\max}(M)$ denote the smallest and the largest characteristic root of a p.d. matrix M. A $p \times p$ diagonal matrix whose diagonal elements are, say, c_1, $c_2, \cdots, c_p$ will be denoted by $D_c(p \times p)$ or simply by D_c. A $p \times p$ unit matrix will be denoted by $I(p)$.

2.1. *Statement and reduction of the problem for the case of Σ and $\Sigma_1\Sigma_2^{-1}$.* We take over from the previous paper [1] the two confidence statements (5.1.5) and (5.2.4) and renumber them as

$$\text{(2.1.1)} \qquad \underset{\sim}{a}'\underset{\sim}{a}\theta_{1\alpha}(p, n) \leqq \underset{\sim}{a}'(D_{1/\sqrt{\theta}}n\Gamma' S\Gamma D_{1/\sqrt{\theta}})\underset{\sim}{a} \leqq \underset{\sim}{a}'\underset{\sim}{a}\theta_{2\alpha}(p, n),$$

$$\text{(2.1.2)} \qquad (n_2/n_1)\theta_{1\alpha}(p, n_1, n_2)\underset{\sim}{b}'S_2\underset{\sim}{b} \leqq \underset{\sim}{b}'(\mu D_{1/\sqrt{\theta}}\mu^{-1}S_1\mu'^{-1}D_{1/\sqrt{\theta}}\mu')\underset{\sim}{b}$$
$$\leqq (n_2/n_1)\theta_{2\alpha}(p, n_1, n_2)\underset{\sim}{b}'S_2\underset{\sim}{b}.$$

These statements are supposed to hold respectively for all nonnull $\underset{\sim}{a}(p \times 1)$ and $\underset{\sim}{b}(p \times 1)$, and each with a confidence coefficient $1 - \alpha$.

In (2.1.1), S stands for the sample dispersion matrix, $n + 1$ for the sample size, and the Θ's for the characteristic roots of Σ. Here Γ is an orthogonal matrix

Received 1/8/54, revised 6/29/54.

given by $\Sigma = \Gamma D_\Theta \Gamma'$, and $\theta_{1\alpha}(p, n)$ and $\theta_{2\alpha}(p, n)$ are subject only to the restriction

$$P(\theta_{1\alpha} \leqq \theta_1 \leqq \theta_p \leqq \theta_{2\alpha} \mid \Sigma) = 1 - \alpha, \tag{2.1.3}$$

where θ_1 and θ_p are the smallest and the largest characteristic roots of nS. Otherwise $\theta_{1\alpha}$ and $\theta_{2\alpha}$ are, for the moment, left flexible, unlike what was done in the previous paper [1].

In (2.1.2), S_1 and S_2 stand for the two sample dispersion matrices, $n_1 + 1$ and $n_2 + 1$ for the two sample sizes, and the Θ's for the characteristic roots of $\Sigma_1\Sigma_2^{-1}$. Here μ is a nonsingular matrix given by $\Sigma_1 = \mu D_\Theta \mu'$ and $\Sigma_2 = \mu\mu'$, while $\theta_{1\alpha}(p, n_1, n_2)$ and $\theta_{2\alpha}(p, n_1, n_2)$ are subject only to the restriction

$$P(\theta_{1\alpha} \leqq \theta_1 \leqq \theta_p \leqq \theta_{2\alpha} \mid \Sigma_1 = \Sigma_2) = 1 - \alpha, \tag{2.1.4}$$

where θ_1 and θ_p are the smallest and the largest characteristic roots of $(n_1/n_2)S_1S_2^{-1}$. Otherwise $\theta_{1\alpha}$ and $\theta_{2\alpha}$ are, for the moment, left free, unlike the development of the previous paper [1].

Let us denote by $c(M)$ any characteristic root of the matrix M. Then it is well known that the statements (2.1.1) and (2.1.2) are respectively equivalent to

$$(1/n)\theta_{1\alpha}(p, n) \leqq \text{all } c(D_{1/\sqrt{\Theta}}\Gamma' S \Gamma D_{1/\sqrt{\Theta}}) \leqq (1/n)\theta_{2\alpha}(p, n), \tag{2.1.5}$$

$$(n_2/n_1)\theta_{1\alpha}(p, n_1, n_2) \leqq \text{all } c(\mu D_{1/\sqrt{\Theta}}\mu^{-1} S_1 \mu'^{-1} D_{1/\sqrt{\Theta}}\mu' S_2^{-1}) \leqq (n_2/n_1)\theta_{2\alpha}(p, n_1, n_2). \tag{2.1.6}$$

We notice that $\Theta_i = c_i(\Sigma)$ in (2.1.5) and $= c_i(\Sigma_1\Sigma_2^{-1})$ in (2.1.6), with $i = 1, \cdots, p$. It is now our purpose to obtain confidence bounds on Θ_i's (or their functions) in terms of $c_i(S)$'s (or their functions) in the case of (2.1.5), and in terms of $c_i(S_1)$ and $c_i(S_2)$ (or their functions) in the case of (2.1.6). For $c_i(\Sigma)$'s the confidence bounds are given by (3.1.3) and (3.1.4), and for $c_i(\Sigma_1\Sigma_2^{-1})$ by (3.2.8).

2.2. *Statement and reduction of the problem for the case of the regression matrix* β. We recall the confidence statement ([1], (6.1.4)), with a confidence coefficient $1 - \alpha$:

$$b - \frac{t_\alpha(n-2)}{\sqrt{n-2}}\sqrt{1-r^2}\,\frac{s_1}{s_2} \leqq \beta \leqq b + \frac{t_\alpha(n-2)}{\sqrt{n-2}}\sqrt{1-r^2}\,\frac{s_1}{s_2}, \tag{2.2.1}$$

where β (which is now a scalar) stands for the population regression of x_1 on x_2 (where x_1 and x_2 have a bivariate normal distribution), b for the sample regression (in a random sample of size $n \geqq 3$), r for the sample correlation, s_1 and s_2 for the two sample standard deviations, and t_α for the upper $\frac{1}{2}\alpha$-point of the t-distribution with D.F. $(n - 2)$.

We also note that

$$b = rs_1/s_2 = rs_1s_2/s_2^2, \qquad \beta = \rho\sigma_1\sigma_2/\sigma_2^2, \tag{2.2.2}$$

where ρ, σ_1, and σ_2 stand respectively for the population correlation coefficient and the two standard deviations.

We now start ([1], Sec. 6.2) with a random sample of size n, with $n > p + q$ and $p \leqq q$, from a $(p + q)$-variate normal population. Next we reduce for the means and set

$$(n-1)\begin{pmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} (Y_1' \;\; Y_2'),$$

where S_{11}, S_{22}, and S_{12} stand respectively for the sample dispersion submatrix of the p-set, that of the q-set, and that between the p-set and the q-set. Here Y_1 and Y_2 have p.d.f. proportional to

$$\exp\left[-\frac{1}{2}\operatorname{tr}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} (Y_1' \;\; Y_2')\right]. \tag{2.2.3}$$

We next recall ([1], Sec. 6.2) that there exist nonsingular $\mu_1(p \times p)$ and $\mu_2(q \times q)$ such that

$$\Sigma_{11}(p \times p) = \mu_1(p \times p)\mu_1'(p \times p), \qquad \Sigma_{22}(q \times q) = \mu_2(q \times q)\mu_2'(q \times q), \tag{2.2.4}$$

$$\Sigma_{12}(p \times q) = \mu_1(p \times p)[D_{\sqrt{\theta}} \;\; 0]\mu_2'(q \times q),$$

where $D_{\sqrt{\theta}}$ is a $p \times p$ matrix and the Θ's are the characteristic roots (all non-negative) of the matrix $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}'$ (i.e., the squares of the population canonical correlations between the p-set and the q-set). As in ([1], Sec. 6.2), we have

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}^{-1} = \left[\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} \begin{pmatrix} I(p) & \vdots & (D_{\sqrt{\theta}} \; 0) \\ \cdots & \vdots & \cdots \\ \begin{pmatrix} D_{\sqrt{\theta}} \\ 0 \end{pmatrix} & \vdots & I(q) \end{pmatrix} \begin{pmatrix} \mu_1' & 0 \\ 0 & \mu_2' \end{pmatrix}\right]^{-1}$$

$$= \left[\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} \begin{pmatrix} D_{\sqrt{1-\theta}} & \vdots & (D_{\sqrt{\theta}} \; 0) \\ \cdots & \vdots & \cdots \\ 0 & \vdots & I(q) \end{pmatrix} \begin{pmatrix} D_{\sqrt{1-\theta}} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ \begin{pmatrix} D_{\sqrt{\theta}} \\ 0 \end{pmatrix} & \vdots & I(q) \end{pmatrix} \begin{pmatrix} \mu_1' & 0 \\ 0 & \mu_2' \end{pmatrix}\right]^{-1} \tag{2.2.5}$$

$$= \begin{pmatrix} \mu_1'^{-1} & 0 \\ 0 & \mu_2'^{-1} \end{pmatrix} \begin{pmatrix} D_{\sqrt{1/(1-\theta)}} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ -\begin{pmatrix} D_{\sqrt{\theta/(1-\theta)}} \\ 0 \end{pmatrix} & \vdots & I(q) \end{pmatrix} \begin{pmatrix} D_{\sqrt{1/(1-\theta)}} & \vdots & -(D_{\sqrt{\theta/(1-\theta)}} \; 0) \\ \cdots & \vdots & \cdots \\ 0 & \vdots & I(q) \end{pmatrix}$$

$$\times \begin{pmatrix} \mu_1^{-1} & 0 \\ 0 & \mu_2^{-1} \end{pmatrix}.$$

Going back to (2.2.3) and using the result that tr $(AB) = $ tr (BA), we have

$$
\begin{aligned}
&\operatorname{tr}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}(Y_1'\ Y_2')\\
&= \operatorname{tr}\left(\begin{array}{c:c} D_{\sqrt{1/(1-\theta)}} & -(D_{\sqrt{\theta/(1-\theta)}}\ 0)\\ \hdashline 0 & I(q)\end{array}\right)\begin{pmatrix}\mu_1^{-1} & 0\\ 0 & \mu_2^{-1}\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}(Y_1'\ Y_2') \qquad (2.2.6)\\
&\qquad\times\begin{pmatrix}\mu_1'^{-1} & 0\\ 0 & \mu_2'^{-1}\end{pmatrix}\left(\begin{array}{c:c} D_{\sqrt{1/(1-\theta)}} & 0\\ \hdashline -\begin{pmatrix}D_{\sqrt{\theta/(1-\theta)}}\\ 0\end{pmatrix} & I(q)\end{array}\right) = \operatorname{tr}\begin{pmatrix}Z_1\\ Z_2\end{pmatrix}(Z_1'\ Z_2'),
\end{aligned}
$$

where

$$Z_1 = \sqrt{1/(1-\theta)}\mu_1^{-1}Y_1 - (D_{\sqrt{\theta/(1-\theta)}}\quad 0)\mu_2^{-1}Y_2\,, \qquad Z_2 = \mu_2^{-1}Y_2\,. \tag{2.2.7}$$

Thus it is easy to check from (2.2.3), (2.2.6), and (2.2.7) that (Z_1, Z_2) have a p.d.f. proportional to

$$\exp\left\{-\tfrac{1}{2}\operatorname{tr}\begin{pmatrix}Z_1\\ Z_2\end{pmatrix}(Z_1'\ Z_2')\right\}. \tag{2.2.8}$$

Consider now, for any two arbitrary nonnull vectors $\underset{\sim}{a}_1(p\times 1)$ and $\underset{\sim}{a}_2(q\times 1)$ and for a fixed positive θ_0, the statement

$$\frac{(\underset{\sim}{a}_1'Z_1Z_2'\underset{\sim}{a}_2)^2}{(\underset{\sim}{a}_1'Z_1Z_1'\underset{\sim}{a}_1)(\underset{\sim}{a}_2'Z_2Z_2'\underset{\sim}{a}_2)} \leqq \theta_0. \tag{2.2.9}$$

This can be written in terms of Y_1 and Y_2 as

$$\frac{[\underset{\sim}{a}_1'\{D_{\sqrt{1/(1-\theta)}}\,\mu_1^{-1}Y_1Y_2'\mu_2'^{-1} - (D_{\sqrt{\theta/(1-\theta)}}\quad 0)\mu_2^{-1}Y_2Y_2'\mu_2'^{-1}\}\underset{\sim}{a}_2]^2}{(\underset{\sim}{a}_2'\mu_2^{-1}Y_2Y_2'\mu_2'^{-1}\underset{\sim}{a}_2)(\underset{\sim}{a}_1'QQ'\underset{\sim}{a}_1)} \leqq \theta_0, \tag{2.2.10}$$

where

$$Q = D_{\sqrt{1/(1-\theta)}}\,\mu_1^{-1}Y_1 - (D_{\sqrt{\theta/(1-\theta)}}\quad 0)\,\mu_2^{-1}Y_2. \tag{2.2.11}$$

Now putting

$$\underset{\sim}{b}_1'(1\times p) = \underset{\sim}{a}_1'D_{\sqrt{1/(1-\theta)}}\mu_1^{-1}, \qquad \underset{\sim}{b}_2'(1\times q) = \underset{\sim}{a}_2'\mu_2^{-1} \tag{2.2.12}$$

and using (2.2.4), we check that (2.2.10) reduces to

$$\frac{[\underset{\sim}{b}_1'(Y_1Y_2' - \beta Y_2Y_2')\underset{\sim}{b}_2]^2}{[\underset{\sim}{b}_2'Y_2Y_2'\underset{\sim}{b}_2][\underset{\sim}{b}_1'(Y_1 - \beta Y_2)(Y_1' - Y_2'\beta')\underset{\sim}{b}_1]} \leqq \theta_0$$

or

$$\frac{[\underset{\sim}{b}_1'(S_{12} - \beta S_{22})\underset{\sim}{b}_2]^2}{(\underset{\sim}{b}_2'S_{22}\underset{\sim}{b}_2)[\underset{\sim}{b}_1'(S_{11} - S_{12}\beta' - \beta S_{12}' + \beta S_{22}\beta')\underset{\sim}{b}_1]} \leqq \theta_0, \tag{2.2.13}$$

where

$$\beta(p \times q) = \mu_1(D_{\sqrt{\theta}} \quad 0)\mu_2^{-1} = \Sigma_{12}\Sigma_{22}^{-1}. \tag{2.2.14}$$

As defined by (2.2.14), β can be appropriately called the matrix of population regression of the p-set on the q-set. It is the only set of population parameters that occurs in statement (2.2.13).

For an $p \times p$ matrix B, let $\mathrm{tr}_s(B)$, for $s = 1, \cdots, p$, stand for the sum of all sth order principal minors of B. It is well known that

$$\mathrm{tr}_s(B) = \sum_{i_1 \neq i_2 \neq \cdots \neq i_s = 1}^{p} c_{i_1}(B)c_{i_2}(B)\cdots c_{i_s}(B), \tag{2.3.1}$$

and, in particular, that $\mathrm{tr}_1(B) = \sum c_i(B) = \sum b_{ii}$, and $\mathrm{tr}_p(B) = \prod c_i(B) = B$. Also well known is that

$$c[A(p \times p)B(p \times p)] = c[B(p \times p)A(p \times p)] \tag{2.3.2}$$

Furthermore we recall

LEMMA A. *The product of two* p.d. *matrices is* p.d. *If $A(p \times q)$ [rank $r \leqq \min(p, q)$] is a matrix with real elements, then AA' is* p.s.d. *of rank r.*

We have also [2] that

$$c_{\min}(A)\, c_{\min}(B) \leqq \text{all } c(AB) \leqq c_{\max}(A)\, c_{\max}(B), \tag{2.3.3}$$

where A and B are two symmetric matrices of which one is p.d. and the other is at least p.s.d. The generalization to the product of a finite number of matrices is obvious [2]. We also take over from [2] the result that

$$c_{\min}(MM') \leqq c^2(M) \leqq c_{\max}(MM'), \tag{2.3.4}$$

where M is a square matrix with real characteristic roots. From (2.3.4) it is easy to see, by replacing A by AB^{-1} (if B is nonsingular), that

$$c_{\min}(AB^{-1})\, c_{\min}(B) \leqq \text{all } c(A) \leqq c_{\max}(AB^{-1})\, c_{\max}(B). \tag{2.3.5}$$

Next, we establish

LEMMA B. *If $d_1 \leqq$ all $c(AB^{-1}) \leqq d_2$, then*

$$(d_1)^t\, \mathrm{tr}_t(B) \leqq \mathrm{tr}_t(A) \leqq (d_2)^t\, \mathrm{tr}_t(B), \qquad t = 1, \cdots, p,$$

where A and B are two $p \times p$ matrices and d_1 and d_2 any two positive numbers such that $d_1 \leqq d_2$.

The conclusion is a necessary (though not a sufficient) condition for the hypothesis.

PROOF. It is easy to check that the statement $d_1 <$ all $c(AB^{-1})$ is equivalent to the statement "A-d_1B is p.d." which again is equivalent to the statement "$A_t - d_1B_t$, for $t = 1, 2, \cdots, p$, is p.d.", where $A_t - d_1B_t$ is a submatrix formed by the intersection of any t rows of $A - d_1B$ with t columns bearing the same numbers. The last statement again is equivalent to the statement $d_1 <$ all $c(A_tB_t^{-1})$.

Now, if all $c(A_t B_t^{-1}) > d_1$, one consequence is that

$$(2.3.6)\quad \prod_{i=1}^{t} c_i(A_t B_t^{-1}) > (d_1)^t, \quad \text{that is,} \quad \frac{|A_t|}{|B_t|} > (d_1)^t, \quad \text{that is,} \quad |A_t| > (d_1)^t\,|B_t|.$$

For a given t, summing over different possible submatrices we have

$$(2.3.7)\qquad \mathrm{tr}_t A > (d_1)^t\,\mathrm{tr}_t B.$$

Using the same kind of argument for the other half of the inequality and remembering that $t = 1, 2, \cdots, p$, and combining, we have the result that

$$(2.3.8)\quad \text{if}\quad d_1 < \text{all } c(AB^{-1}) < d_2, \quad \text{then}\quad (d_1)^t\,\mathrm{tr}_t(B) < \mathrm{tr}_t(A) < (d_2)^t\,\mathrm{tr}_t(B), \qquad t = 1, \cdots, p.$$

By a slight rephrasing (which is obviously permissible here) we can obtain Lemma B) from (2.3.8). We recall the three following well known lemmas, repeatedly used in [2].

LEMMA C. *The statement "$g_1 \leqq$ all $c(M) \leqq g_2$ (for a $p \times p$ real matrix M with real roots)" is equivalent to the statement "$g_1 \leqq \underline{d}'(1 \times p)M(p \times p)\underline{d}(p \times 1) \leqq g_2$ (for all arbitrary vectors $\underline{d}$ of unit modulus)".*

LEMMA D. *The statement "$g_1 \leqq$ all $c(M_1 M_2^{-1}) \leqq g_2$ (for two $p \times p$ real matrices M_1 and M_2 with real roots, M_2 being nonsingular" is equivalent to the statement "$g_1 \leqq \underline{d}'(1 \times p)M_1(p \times p)\underline{d}(p \times 1) \,/\, \underline{d}'(1 \times p)M_2(p \times p)\underline{d}(p \times 1) \leqq g_2$ (for all arbitrary nonnull vectors $\underline{d}$)".*

LEMMA E. *The statement "$\underline{x}'(1 \times q)\underline{x}(q \times 1) \leqq h\ (> 0)$" is equivalent to the statement "$|\underline{x}'(1 \times q)\underline{d}(q \times 1)| \leqq \sqrt{h}$ (for all arbitrary vectors $\underline{d}$ of unit modulus)".*

2.4. *A result in set-theoretic logic.* It is well known that the statement "If E_1, then E_2" is equivalent to the statement "E_2 is a necessary condition for E_1" which again is equivalent to the statement "$E_1 \subset E_2$". All these statements imply that "$P(E_1) \leqq P(E_2)$", which is a necessary (though not a sufficient) condition for the other statements. This will be used in the derivation of the confidence bounds.

3. Confidence bounds on $c(\Sigma)$'s and $c(\Sigma_1\Sigma_2^{-1})$'s.

3.1. *Bounds on $c(\Sigma)$'s.* Starting from (2.1.5) and noting that

$$(3.1.1)\qquad c(D_{1/\sqrt{\theta}}\Gamma' S \Gamma D_{1/\sqrt{\theta}}) = c(S\Gamma D_{1/\theta}\Gamma') = c(S\Sigma^{-1}),$$

we have, with a confidence coefficient $1 - \alpha$, the equivalent confidence bounds

$$(3.1.2)\qquad \begin{aligned} (1/n)\theta_{1\alpha}(p, n) &\leqq \text{all } c(S\Sigma^{-1}) \leqq (1/n)\theta_{2\alpha}(p, n), \\ n\theta_{1\alpha}^{-1}(p, n) &\geqq \text{all } c(\Sigma S^{-1}) \geqq n\theta_{2\alpha}^{-1}(p, n). \end{aligned}$$

From (2.3.6) we observe that this implies

$$(3.1.3)\qquad n\theta_{1\alpha}^{-1}(p, n)C_{\max}(S) \geqq \text{all } c(\Sigma) \geqq n\theta_{2\alpha}^{-1}(p, n)c_{\min}(S),$$

which is thus a set of simultaneous confidence bounds with a confidence coefficient $\geqq 1 - \alpha$. We note that, by using Lemma C, we can replace "all $c(\Sigma)$" occurring in the middle of (3.1.3) by "$\underset{\sim}{a}'\Sigma\underset{\sim}{a}$ (for all arbitrary vectors $\underset{\sim}{a}$ of unit modulus)."

From Lemma B we also observe that (3.1.2) implies

$$(3.1.4) \qquad [n\theta_{1\alpha}^{-1}(p, n)]^t \operatorname{tr}_t(S) \geqq \operatorname{tr}_t(\Sigma) \geqq [n\theta_{2\alpha}^{-1}(p, n)]^t \operatorname{tr}_t(S),$$

for $t = 1, 2, \cdots, p$, which is thus also another set of simultaneous confidence bounds with a confidence coefficient $\geqq 1 - \alpha$. Using (2.3.1), $\operatorname{tr}_t(S)$ and $\operatorname{tr}_t(\Sigma)$ are easily calculated in terms of θ_i's and Θ_i's.

3.2. *Bounds on $c(\Sigma_1\Sigma_2^{-1})$'s.* Starting from (2.1.6) we have, with a confidence coefficient $1 - \alpha$, the confidence bounds

$$(3.2.1) \qquad (n_1/n_2)\theta_{1\alpha}^{-1}(p, n_1, n_2) \geqq \text{all } c(S_2(\mu')^{-1}D_{\sqrt{\Theta}}\mu' S_1^{-1}\mu D_{\sqrt{\Theta}}\mu^{-1}) \geqq (n_1/n_2)\theta_{2\alpha}^{-1}(p, n_1, n_2).$$

Using (2.3.2) and (2.3.6) we have

$$(3.2.2) \qquad c_{\max}[S_2(\mu')^{-1}D_{\sqrt{\Theta}}\mu' S_1^{-1}\mu D_{\sqrt{\Theta}}\mu^{-1}]c_{\max}(S_2^{-1}) \geqq \text{all } c[(\mu')^{-1}D_{\sqrt{\Theta}}\mu' S_1^{-1}\mu D_{\sqrt{\Theta}}\mu^{-1}] = \text{all } c(S_1^{-1}\Delta) \geqq c_{\min}[S_2(\mu')^{-1}D_{\sqrt{\Theta}}\mu' S_1^{-1}\mu D_{\sqrt{\Theta}}\mu^{-1}]\, c_{\min}(S_2^{-1}),$$

where

$$(3.2.3) \qquad \Delta = (\mu D_{\sqrt{\Theta}}\mu^{-1})((\mu')^{-1}D_{\sqrt{\Theta}}\mu') = (\mu D_{\sqrt{\Theta}}\mu^{-1})(\mu D_{\sqrt{\Theta}}\mu^{-1})'.$$

In the same way we have

$$(3.2.4) \qquad c_{\max}(S_1^{-1}\Delta)\, c_{\max}(S_1) \geqq \text{all } c(\Delta) \geqq c_{\min}(S_1^{-1}\Delta)\, c_{\min}(S_1).$$

Furthermore, noting that

$$(3.2.5) \qquad c(\mu D_{\sqrt{\Theta}}\mu^{-1}) = c(D_{\sqrt{\Theta}}) = \sqrt{\Theta} = c(\mu'^{-1}\, D_{\sqrt{\Theta}}\mu'),$$

and using (2.3.5), we have

$$(3.2.6) \qquad c_{\max}(\Delta) \geqq \text{all } c^2(\mu D_{\sqrt{\Theta}}\mu^{-1}) = \text{all } c^2(D_{\sqrt{\Theta}}) = \text{all } \Theta_i \geqq c_{\min}(\Delta).$$

Combining (3.2.2), (3.2.4) and (3.2.6), we have

$$(3.2.7) \qquad c_{\max}(S_2(\mu')^{-1}D_{\sqrt{\Theta}}\mu' S_1^{-1}\mu D_{\sqrt{\Theta}}\mu^{-1})\, c_{\max}(S_2^{-1})\, c_{\max}(S_1) \geqq \text{all } \Theta_i \geqq c_{\min}(S_2(\mu')^{-1}D_{\sqrt{\Theta}}\mu' S_1^{-1}\mu D_{\sqrt{\Theta}}\mu^{-1})\, c_{\min}(S_2^{-1})\, c_{\min}(S_1).$$

From this it is easy to check that (3.2.1) implies

$$(3.2.8) \qquad (n_1/n_2)\Theta_{1\alpha}^{-1}(p, n_1, n_2)\, c_{\max}(S_2^{-1})\, c_{\max}(S_1) \geqq \text{all } c(\Sigma_1\Sigma_2^{-1}) \geqq (n_1/n_2)\theta_{2\alpha}^{-1}(p, n_1, n_2)\, c_{\min}(S_2^{-1})\, c_{\min}(S_1),$$

which is thus a set of simultaneous confidence bounds with a confidence coefficient $\geqq 1 - \alpha$. We observe that, by using Lemma D we can replace "all $c(\Sigma_1\Sigma_2^{-1})$" occurring in the middle of (3.2.8) by "$\underset{\sim}{a}'\Sigma_1\underset{\sim}{a}/\underset{\sim}{a}'\Sigma_2\underset{\sim}{a}$ (for all arbitrary nonnull vectors $\underset{\sim}{a}(p \times 1)$)." We notice that

$$c_{\max}(S_2^{-1}) = 1/c_{\min}(S_2), \qquad c_{\min}(S_2^{-1}) = 1/c_{\max}(S_2).$$

Confidence bounds in terms of tr_t could also be given as in (3.1.4), but in this case the bounds would be more complicated and would appear to be less worthwhile than in the previous case.

3.3. *Determination of the constants* $(\theta_{1\alpha}(p, n), \theta_{2\alpha}(p, n))$ *and* $(\theta_\alpha(p, n_1, n_2), \theta_{2\alpha}(p, n_1, n_2))$ *occurring in the confidence bounds.* It has been stated in Section 2 that the pair $\theta_{1\alpha}(p, n)$, $\theta_{2\alpha}(p, n)$ for the first problem and the pair $\theta_{1\alpha}(p, n_1, n_2)$, $\theta_{2\alpha}(p, n_1, n_2)$ for the second problem satisfy respectively the conditions (2.1.3) and (2.1.4), but are otherwise free. It is well known how the shortness (in the sense of probability) of a confidence interval (or intervals) ties in with the power of the associated test. Let us consider the associated tests, or rather, the acceptance regions of the respective hypotheses (i) $H(\Sigma = \Sigma_0)$ and (ii) $H(\Sigma_1 = \Sigma_2)$. They are, respectively,

$$H(\Sigma = \Sigma_0): \theta_{1\alpha}(p, n) \leqq \theta_1 \leqq \theta_p \leqq \theta_{2\alpha}(p, n), \tag{3.3.1}$$

$$H(\Sigma_1 = \Sigma_2): \theta_{1\alpha}(p, n_1, n_2) \leqq \theta_1 \leqq \theta_p \leqq \theta_{2\alpha}(p, n_1, n_2). \tag{3.3.2}$$

In the first case it is possible to choose $\theta_{1\alpha}$ and $\theta_{2\alpha}$ (and this choice will be unique) so as to let the second kind of error (which, aside from p, n and α, depends only on the characteristic roots of $\Sigma\Sigma_0^{-1}$) have a (local) minimum, that is to let the power have a local maximum at $\Sigma = \Sigma_0$, when $\Sigma \neq \Sigma_0$ is supposed to be the alternative. In this case it so happens that the resulting power function then monotonically increases as each $c_i(\Sigma\Sigma_0^{-1})$ tends away from unity, provided that all are $\geqq 1$ or $\leqq 1$, to begin with.

In the second case, we have an exactly similar situation, $H(\Sigma = \Sigma_0)$ being replaced by $H(\Sigma_1 = \Sigma_2)$ and $\Sigma\Sigma_0^{-1}$ being replaced by $\Sigma_1\Sigma_2^{-1}$. The effect of this on the shortness, in the probability sense, of the resulting confidence bounds is obvious and need not be discussed in detail.

The results just stated are proved in another paper to be published shortly. It may be noticed, however, that for any pair $(\theta_{1\alpha}, \theta_{2\alpha})$ subject only to (2.1.3) or (2.1.4), we are going to get anyway the confidence bounds of Sections 3.1 and 3.2, with confidence coefficients $\geqq 1 - \alpha$, the only difference being that they will not have the property of "shortness" possessed by those that are based on $(\theta_{1\alpha}, \theta_{2\alpha})$ determined in the above way.

4. Confidence bounds on the regression matrix $\Sigma_{12}\Sigma_{22}^{-1}$ or β. It is well known [1] that the statement (2.2.13), for all arbitrary nonnull $\underset{\sim}{b}_1$ and $\underset{\sim}{b}_2$, is exactly equivalent to

$$\text{all } \theta_i \leqq \theta_0 \qquad \text{or} \qquad \theta_p \leqq \theta_0, \tag{4.1}$$

where the θ_i's, for $i = 1, \cdots, p$ and $0 \leqq \theta_1 \leqq \cdots \leqq \theta_p \leqq 1$, are the roots of the determinantal equation in θ

$$(4.2)\quad |\theta(S_{11} - S_{12}\beta' - \beta S_{12}' + \beta S_{22}\beta') - (S_{12} - \beta S_{22})S_{22}^{-1}(S_{12}' - S_{22}\beta')| = 0$$

Now put $\lambda \equiv \theta/(1 - \theta)$, so that we have, from (4.2), the determinantal equation in λ

$$(4.3)\quad |\lambda(S_{11} - S_{12}S_{22}^{-1}S_{12}') - (S_{12}S_{22}^{-1} - \beta)S_{22}(S_{22}^{-1}S_{12}' - \beta')| = 0.$$

Statement (4.1) can now be replaced by the statement that the largest root of (4.3) is not greater than $\lambda \equiv \theta_0/(1 - \theta_0)$, that is,

$$(4.4)\quad \text{all } c[(S_{11} - S_{12}S_{22}^{-1}S_{12}')^{-1}(B - \beta)S_{22}(B' - \beta')] \leqq \theta_0/(1 - \theta_0),$$

where $B(p \times q) = S_{12}S_{22}^{-1}$. This B may be called appropriately the matrix of sample regression of the p-set on the q-set.

We note that (4.4) is equivalent to (4.1) which again is equivalent to (2.2.9), so that θ_p is the largest characteristic root of the matrix $(Z_1Z_1')^{-1}(Z_1Z_2')(Z_2Z_2')^{-1} \times (Z_2Z_1')$, where (Z_1, Z_2) have the p.d.f. given by (2.2.8). The joint distribution of these central θ_i's, and also of the largest root θ_p are known; thus all that we have to do to make (4.4), that is (4.1), that is, (2.2.9), a simultaneous confidence statement with a joint confidence coefficient $1 - \alpha$ is to choose $\theta_0 = \theta_\alpha(p, q, n - 1) = \theta_\alpha$ (say), where θ_0 or θ_α is defined by $P(\text{central } \theta_p \geqq \theta_0) = \alpha$.

Now, as in Sections 3.1 and 3.2, using (2.3.5) and the result in Section 2.4, we have from (4.4), with a joint confidence coefficient $\geqq 1 - \alpha$, the simultaneous confidence statement that

$$(4.5)\quad \text{all } c[(B - \beta)(B' - \beta')] \leqq [\theta_\alpha/(1 - \theta_\alpha)]c_{\max}(S_{11} - S_{12}S_{22}^{-1}S_{12}')c_{\max}(S_{22}^{-1}).$$

We now note that

$$c_{\max}(S_{22}^{-1}) = 1/c_{\min}(S_{22}),$$

$$c_{\max}(S_{11} - S_{12}S_{22}^{-1}S_{12}') \leqq c_{\max}(S_{11})c_{\max}(I - S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}'),$$

$$c_{\max}(I - S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}') = 1 - c_{\min}(S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}').$$

Using these, we check that (4.5) can be replaced (with a confidence coefficient $\geqq 1 - \alpha$) by

$$(4.6)\quad \text{all } c[(B - \beta)(B' - \beta')] \leqq \frac{\theta_\alpha}{1 - \theta_\alpha}[1 - c_{\min}(S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}')]\frac{c_{\max}(S_{11})}{c_{\min}(S_{22})}.$$

Letting h denote the right side of (4.6), and applying the Lemmas C and E to (4.6) we have, with a joint confidence coefficient $\geqq 1 - \alpha$, the following equivalent simultaneous confidence statements for all arbitrary unit modulus vectors $\underset{\sim}{d}_1(p \times 1)$ and $\underset{\sim}{d}_2(q \times 1)$,

$$(4.7)\quad |\underset{\sim}{d}_1'(B - \beta)\underset{\sim}{d}_2| \leqq \sqrt{h}, \qquad \underset{\sim}{d}_1'B\underset{\sim}{d}_2 - \sqrt{h} \leqq \underset{\sim}{d}_1'\beta\underset{\sim}{d}_2 \leqq \underset{\sim}{d}_1'B\underset{\sim}{d}_2 + \sqrt{h}.$$

A set of simultaneous confidence bounds on just the elements β_{ij} of the β-matrix would be a subset of the bounds on the total set $\underline{d}_1'\beta\underline{d}_2$. It is worthwhile to check that, if $p = q = 1$, (4.7) reduces, as it should, to (2.2.1). Also, if $p = 1$, we should have another special case of (4.7) giving a set of simultaneous confidence bounds on all linear functions of the partial regressions of one variate on several others. Thus, in several ways, (4.7) seems to be an appropriate generalization of (2.2.1).

REFERENCES

[1] S. N. Roy and R. C. Rose, "Simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 513–536.

[2] S. N. Roy, "A useful theorem in matrix theory," *Proc. Amer. Math. Soc.*, Vol. 5 (1954), pp. 635–638.

TABLES FOR THE DISTRIBUTION OF THE NUMBER OF EXCEEDANCES[1]

BY BENJAMIN EPSTEIN

Wayne University

1. Introduction and Summary. Consider a random sample of size n taken from a continuous distribution $f(x)$. Let another random sample, independent of the first sample and also of size n, be drawn from the same population. Let U_r^n be the random variable associated with the number of values in the second sample which exceed the rth smallest value in the first sample. Similarly let V_s^n be the random variable associated with the number of values in the second sample which exceed the sth largest value in the first sample. Due to the fact that the rth smallest value in a sample of size n is at the same time the sth largest value in the sample with $s = n - r + 1$, it follows that

$$\text{(1)} \qquad \Pr(U_r^n = x) = \Pr(V_s^n = x),$$

$$s = n - r + 1; \quad r = 1, 2, \cdots, n; \quad x = 0, 1, 2, \cdots, n.$$

The probability distribution of U_r^n (and hence of V_s^n) is given by:

$$\text{(2)} \quad \Pr(U_r^n = x) = \binom{n-x+r-1}{r-1}\binom{n-r+x}{x}\Big/\binom{2n}{n} = \tfrac{1}{2}P_{n-x+r-1,r-1}\,P_{n-r+x,x}/P_{2n,n},$$

$$x = 0, 1, 2, \cdots, n.$$

Formula (2) can be proved by combinatorial methods; details are omitted. An alternative formula, derived in another way [3], is

$$\text{(2a)} \qquad \Pr(U_r^n = x) = \tfrac{1}{2}\binom{n-1}{r-1}\binom{n}{x}\Big/\binom{2n-1}{n-r+x} = \tfrac{1}{2}P_{n-1,r-1}P_{n,x}/P_{2n-1,n-r+x}.$$

In formulae (2) and (2a), $P_{n,x} = (\tfrac{1}{2})^n\binom{n}{x}$. Formulae in terms of $P_{n,x}$ are particularly convenient for hand computation, since one can use the extensive tables of the binomial probability distribution published by the National Bureau of Standards.

If the values of $\Pr(U_r^n \leqq x)$, for $x = 0, 1, 2, \cdots, n - 1, r = 1, 2, \cdots, n$ are written (for fixed n) in matrix form, one notes certain useful symmetries, which can be expressed by the identities

$$\text{(3)} \qquad \Pr(U_r^n \leqq x) = \Pr(U_{x+1}^n \leqq r - 1),$$

$$\text{(4)} \qquad \Pr(U_r^n \leqq x) + \Pr(U_{n-r+1}^n \leqq n - x - 1) = 1.$$

If one takes $x = n - r$ in (4) and uses the relation (3), it is readily verified that

$$\text{(5)} \qquad \Pr(U_r^n \leqq n - r) = \tfrac{1}{2}.$$

Received 7/20/53, revised 9/18/53.

[1] Work sponsored by the Office of Naval Research.

TABLE 1
$\Pr(U_r^n \leqq x)$

n	r	$x = 0$	1	2	3	4	5	6	7	8	9	10	11	12	13
2	1	.1667													
3	1	.0500	.2000												
4	1	.0143	.0714	.2143											
	2		.2429												
5	1	$.0^{2}397$	.0238	.0833	.2222										
	2		.1032	.2619											
6	1	$.0^{2}108$	$.0^{2}758$	.0303	.0909	.2273									
	2		.0400	.1212	.2727										
	3			.2835											
7	1	$.0^{3}291$	$.0^{2}233$	.0105	.0350	.0962	.2308								
	2		.0146	.0513	.1329	.2797									
	3			.1431	.2960										
8	1	$.0^{4}777$	$.0^{3}699$	$.0^{2}350$	.0128	.0385	.1000	.2333							
	2		$.0^{2}505$	.0203	.0594	.1410	.2846								
	3			.0660	.1573	.3042									
	4				.3096										
9	1	$.0^{4}206$	$.0^{3}206$	$.0^{2}113$	$.0^{2}452$	.0147	.0412	.1029	.2353						
	2		$.0^{2}169$	$.0^{2}761$	.0249	.0656	.1471	.2882							
	3			.0283	.0767	.1674	.3100								
	4				.1735	.3186									
10	1	$.0^{5}541$	$.0^{4}595$	$.0^{3}357$	$.0^{2}155$	$.0^{2}542$	.0163	.0433	.1053	.2368					
	2		$.0^{3}547$	$.0^{2}274$	$.0^{2}988$	.0286	.0704	.1517	.2910						
	3			.0115	.0349	.0849	.1749	.3142							
	4				.0894	.1849	.3250								
	5					.3281									

TABLE 1—*Concluded*

11	1	$.0^{5}142$	$.0^{4}170$	$.0^{3}111$	$.0^{3}516$	$.0^{2}193$	$.0^{2}619$	.0175	.0451	.1071	.2381				
	2		$.0^{3}173$	$.0^{3}953$	$.0^{2}376$	.0119	.0317	.0743	.1554	.2932					
	3			$.0^{2}446$	.0150	.0402	.0913	.1807	.3176						
	4				.0431	.0992	.1935	.3297							
	5					.1974	.3350								
12	1	$.0^{6}370$	$.0^{5}481$	$.0^{4}337$	$.0^{3}168$	$.0^{3}673$	$.0^{2}229$	$.0^{2}686$	.0186	.0466	.1087	.2391			
	2		$.0^{4}536$	$.0^{3}322$	$.0^{2}138$	$.0^{2}471$	.0136	.0343	.0775	.1584	.2950				
	3			$.0^{2}166$	$.0^{2}614$	.0180	.0447	.0965	.1854	.3202					
	4				.0196	.0498	.1069	.2002	.3334						
	5					.1102	.2068	.3401							
	6						.3421								
13	1	$.0^{7}961$	$.0^{5}135$	$.0^{4}101$	$.0^{4}538$	$.0^{3}229$	$.0^{3}824$	$.0^{2}261$	$.0^{2}745$	.0196	.0478	.1100	.2400		
	2		$.0^{4}163$	$.0^{3}106$	$.0^{3}491$	$.0^{2}180$	$.0^{2}558$	.0151	.0365	.0801	.1609	.2965			
	3			$.0^{3}601$	$.0^{2}242$	$.0^{2}771$	.0207	.0484	.1008	.1891	.3224				
	4				$.0^{2}847$	.0236	.0554	.1131	.2055	.3364					
	5					.0576	.1189	.2142	.3441						
	6						.2169	.3475							
14	1	$.0^{7}249$	$.0^{6}374$	$.0^{5}299$	$.0^{4}170$	$.0^{4}763$	$.0^{3}290$	$.0^{3}966$	$.0^{2}290$	$.0^{2}797$	.0204	.0489	.1111	.2407	
	2		$.0^{5}491$	$.0^{4}344$	$.0^{3}171$	$.0^{3}670$	$.0^{2}221$	$.0^{2}638$	.0164	.0384	.0824	.1630	.2978		
	3			$.0^{3}211$	$.0^{3}919$	$.0^{2}316$	$.0^{2}915$	.0230	.0516	.1043	.1923	.3242			
	4				$.0^{2}351$	.0107	.0271	.0601	.1182	.2099	.3388				
	5					.0285	.0642	.1259	.2200	.3473					
	6						.1284	.2247	.3518						
	7							.3532							
15	1	$.0^{8}645$	$.0^{6}103$	$.0^{6}887$	$.0^{5}526$	$.0^{4}250$	$.0^{3}100$	$.0^{3}350$	$.0^{2}110$	$.0^{2}316$	$.0^{2}843$	.0211	.0498	.1121	.2414
	2		$.0^{5}146$	$.0^{4}109$	$.0^{4}579$	$.0^{3}242$	$.0^{3}850$	$.0^{2}260$	$.0^{2}710$	.0176	.0400	.0843	.1648	.2989	
	3			$.0^{4}725$	$.0^{3}339$	$.0^{2}125$	$.0^{2}389$	.0105	.0251	.0543	.1074	.1949	.3257		
	4				$.0^{2}141$	$.0^{2}461$	.0127	.0302	.0641	.1225	.2135	.3408			
	5					.0134	.0328	.0697	.1318	.2249	.3499				
	6						.0716	.1362	.2311	.3552					
	7							.2331	.3576						

		$x = 0$	1	2	3	4	5	6	7	8	9
20	1	$.0^{11}725$	$.0^{9}152$	$.0^{8}168$	$.0^{7}128$	$.0^{7}771$	$.0^{6}385$	$.0^{5}167$	$.0^{5}644$	$.0^{4}225$	$.0^{4}727$
	2		$.0^{9}291$	$.0^{7}291$	$.0^{6}203$	$.0^{5}110$	$.0^{5}501$	$.0^{4}197$	$.0^{4}685$	$.0^{3}216$	$.0^{3}624$
	3			$.0^{6}265$	$.0^{5}168$	$.0^{5}832$	$.0^{4}343$	$.0^{3}122$	$.0^{3}386$	$.0^{2}110$	$.0^{2}287$
	4				$.0^{5}969$	$.0^{4}438$	$.0^{3}164$	$.0^{3}532$	$.0^{2}153$	$.0^{2}396$	$.0^{2}935$
	5					$.0^{3}180$	$.0^{3}616$	$.0^{2}182$	$.0^{2}477$	.0112	.0242
	6						$.0^{2}192$	$.0^{2}519$	.0124	.0268	.0527
	7							.0128	.0281	.0555	.1001
	8								.0564	.1025	.1703
	9									.1715	.2636
	10										.3762
		$x = 10$	11	12	13	14	15	16	17	18	
20	1	$.0^{3}218$	$.0^{3}614$	$.0^{2}164$	$.0^{2}416$	.0101	.0236	.0530	.1154	.2436	
	2	$.0^{3}167$	$.0^{2}418$	$.0^{2}983$	.0218	.0457	.0909	.1708	.3025		
	3	$.0^{2}691$	.0155	.0324	.0637	.1176	.2037	.3307			
	4	.0204	.0412	.0776	.1367	.2253	.3474				
	5	.0479	.0880	.1504	.2401	.3582					
	6	.0954	.1601	.2503	.3655						
	7	.1666	.2572	.3705							
	8	.2616	.3738								
	9	.3756									
	10										

TABLE 2

Values of $\Pr(U_r^5 \leqq x)$

r	$x = 0$	1	2	3	4
1	$.0^2397$	.0238	.0833	.2222	.5000
2	.0238	.1032	.2619	.5000	.7778
3	.0833	.2619	.5000	.7381	.9167
4	.2222	.5000	.7381	.8968	.9762
5	.5000	.7778	.9167	.9762	$.9^2603$

Proofs[2] of (3), (4), and (5) can be obtained by using the results of pages 257–258 of [3]. Because of these symmetries, the complete matrix (for any fixed r) can be constructed if one knows only the quantities, $\Pr(U_r^n \leqq x)$, $r = 1(1)[n/2]$, $x = r-1, r, r+1, \cdots, n-r-1$. In Table 1 these values are given[3] for $n = 2(1)15(5)20$. To see how the complete matrix is obtained from Table 1, it is interesting to verify, using (3), (4), and (5), that the complete matrix, in the special case $n = 5$, is given by Table 2.

A somewhat different, but related, exceedance problem is to take two random samples of size n from a continuous distribution $f(x)$. Let us for convenience attach the letter x to one of the samples and the letter y to the other sample. Further let $x_{r,n}$ and $y_{r,n}$ be respectively the rth smallest observations in each of the samples. Let us define $z_{r,n} = \max(x_{r,n}, y_{r,n})$. If $z_{r,n} = x_{r,n}$, count the number of y's which are $\geqq x_{r,n}$; if $z_{r,n} = y_{r,n}$, count the number of x's which are $\geqq y_{r,n}$. Denoting the number of exceedances as W_r^n, it is readily seen from (1) that the probability distribution of W_r^n is given by

$$(6) \quad \Pr(W_r^n = x) = 2\binom{n-x+r-1}{r-1}\binom{n-r+x}{x} \Big/ \binom{2n}{n}, \qquad x = 0, 1, 2, \cdots, n-r.$$

It is evident from the definition that,

$$(7) \qquad \Pr(W_r^n \leqq x) = 1, \qquad x \geqq n-r.$$

Clearly one can find the values of $\Pr(W_r^n \leqq x)$ by using Table 1. Thus, for example, in the special case $n = 5$ one obtains Table 3.

2. Applications of exceedance theory. There are three principal uses of exceedance theory. These are:

(a) *Floods and droughts.* This theory was used by H. A. Thomas, Jr. [6] in making predictions about the recurrences of floods and droughts in the future on the basis of what is known from past data. In recent papers by Chow [1], [2], the interested reader will find further work in this direction.

[2] We wish to acknowledge with thanks a communication from Dr. E. J. Gumbel on this point.

[3] In Wayne University Technical Report No. 6 (July 1953) values were given for $n = 2(1)20(5)50$. We have also considered the practically important case where the two samples may be of unequal size. Tables for selected pairs of unequal values of the sample size will be available in the near future.

TABLE 3

$\Pr(W_r^5 \leqq x)$

r	$x = 0$	1	2	3	4
1	$.0^2794$	.0476	.1667	.4444	1.0000
2	.0476	.2064	.5238	1.0000	
3	.1667	.5238	1.0000		
4	.4444	1.0000			
5	1.0000				

(b) *Non-parametric tests for slippage.* The functions U_r^n, V_s^n, and W_r^n can be used to give two-sample nonparametric tests for slippage of the mean. There are close connections between the results in this paper and recent tests for slippage by Mosteller and Tukey [4] and [5].

(c) *Life testing.* It is a characteristic feature of life tests that data become available in order of size. Thus it becomes very natural to apply exceedance theory, which is based purely on order statistics. By so doing it is possible in many cases to shorten both the average time and average number of items destroyed in order to reach a decision as to whether or not the items in one population are in some sense superior to the items in another population.

3. Numerical examples.

EXAMPLE 1. What is the probability that the third largest flood during the past 20 years will be exceeded at least once during the next 20 years? *Answer.* The probability is

$$p = 1 - \Pr(V_3^{20} = 0) = 1 - \Pr(U_{18}^{20} = 0) = 1 - .1154 = .8846.$$

EXAMPLE 2. During a period of 20 years the lowest observed annual rainfall in a certain locality was 8.6 inches. What is the probability that in the next 20 years at least two of the years will have rainfall $\leqq$ 8.6 inches? *Answer.* The probability is $p = \Pr(U_1^{20} \leqq 18) = .2436$.

EXAMPLE 3. (one-sided test): We are now interested in making a choice between two lots A and B. In particular we are interested in some characteristic such as life or strength, where data become available in order of magnitude. Let it be known a priori that the probability density function associated with lot B is either the same as that of lot A or is displaced to the left (e.g., is inferior). Put in another way, we are thinking of a case where the only relevant parameter is some measure of slippage. We wish to test the hypothesis H_0 of no displacement against the alternative H_1 that B is displaced to the left of A. The Type I error is taken to be $\leqq .05$. Ten items are drawn from each of the lots and placed on life test. It is decided in advance that a decision will be based on how many failures occur in the sample from B before the second failure occurs in the sample from A. The two samples are put on test simultaneously and give the pattern $bbbabbb\cdots$, where a denotes a failure in the sample drawn from A, b denotes a

failure in the sample drawn from B. The experiment is stopped at the seventh failure with rejection of the null hypothesis, because Pr $(U_2^{10} \leqq 4) = .0286 < .05$. If, however, we had obtained a pattern like *babba* $\cdots$, we would have stopped experimentation after the fifth failure with the acceptance of H_0.

EXAMPLE 4. (two-sided test): Given two lots A and B, we wish to test the null hypothesis that the life distributions of A and B are the same against the alternative that they are different. As in Example 3, let 10 items be drawn at random from each of the two lots and placed on life test. It is decided in advance that our decision will be based on the statistic W_2^{10}. If, for example, the failure pattern observed is *aaaaabaa* $\cdots$, the experiment will be terminated on the eighth trial with rejection of the null hypothesis (on the .05 level of significance). This is because Pr $(W_2^{10} \leqq 3) = .0198$. On the other hand a pattern like *babba* $\cdots$ would lead to acceptance of the null hypothesis on the fifth trial.

4. Discussion. Fairly extensive random sampling experiments have shown that the statistics W_1^{10}, W_2^{10}, and W_3^{10} are more effective than the run test, and somewhat less effective than the Wilcoxon rank test, for detecting slippage of the mean in the case where the underlying distributions are normal, all with the same variance. Since the improvement in power obtained by using W_2^{10} or W_3^{10} rather than W_1^{10} is minor in this case, there are sound practical reasons for preferring W_1^{10}. Decisions based on this statistic can be made at a great saving in average time to decision, as well as average number of items destroyed. It should be noted in Example 4 that if decisions were based on W_1^{10}, we would have truncated testing on the fifth trial with the rejection of H_0, since Pr $(W_1^{10} \leqq 5) = .0325$.

A detailed discussion of the points raised in the last paragraph will appear elsewhere.

5. Acknowledgement. I wish to thank John Lay for his work in computing the tables.

REFERENCES

[1] VEN TE CHOW, "A general formula for hydrologic frequency analysis," *Trans. Amer. Geophys. Union*, Vol. 32, No. 2 (1951), pp. 231–282.

[2] VEN TE CHOW, *Frequency Analysis of Hydrologic Data with Special Application to Rainfall Intensities*, University of Illinois Bulletin, Vol. 50, No. 81, July, 1953.

[3] E. J. GUMBEL AND H. VON SCHELLING, "The distribution of the number of exceedances," *Ann. Math. Stat.*, Vol. 21, (1952), pp. 247–262.

[4] F. MOSTELLER, "A k-sample slippage test for an extreme population," *Ann. Math. Stat.*, Vol. 19, (1948), pp. 58–65.

[5] F. MOSTELLER AND J. W. TUKEY, "Significance levels for a k-sample slippage test," *Ann. Math. Stat.*, Vol. 21, (1950), pp. 120–123.

[6] H. A. THOMAS, JR., "Frequency of minor floods," *J. Boston Soc. Civil Engineers*, Vol. 35, (1948), pp. 425–442.

ON THE FACTORIZATION OF DISTRIBUTIONS

By Henry Teicher

Purdue University

1. Summary. A family of probability distributions is called "factor-closed" (f.c.) if it is closed under the operation of factorization. The classical binomial family and certain generalizations of it are shown to be f.c. The multinomial family is also f.c. Most families of infinitely divisible distributions are not f.c.

2. Introduction. If $F_1(x)$ and $F_2(x)$ are any two cumulative distribution functions (c.d.f.'s), the convolution (denoted by $*$) of F_1 with F_2 is again a c.d.f. say

$$(1) \qquad F = F(x) = F_1 * F_2 = \int_{-\infty}^{\infty} F_1(x-y)\, dF_2(y) = \Pr\{X < x\}$$

where Pr denotes probability measure and X is called the random variable (r.v.) possessing the c.d.f. F. Further, if X_1 and X_2 are independent r.v.'s having c.d.f.'s F_1 and F_2 with corresponding Fourier transforms or characteristic functions (c.f.'s) $\phi_{x_1}(t)$ and $\phi_{x_2}(t)$, then $F = F_1 * F_2$ is the c.d.f. of $X = X_1 + X_2$ having, as is well known, the c.f.

$$(2) \qquad \phi_x(t) = \phi_{x_1}(t) \cdot \phi_{x_2}(t) = \int_{\infty}^{\infty} e^{itx}\, dF(x).$$

If one commences with $\phi_x(t)$ or $F(x)$, any such representation as (2) or (1) is termed a *factorization* of $\phi_x(t)$ or $F(x)$ and the components $\phi_{x_i}(t)$ or $F_i(x)$ are called *factors*.

For an arbitrary distribution F, factorization is not unique. That is, $F = F_1 * F_2 = F_1 * F_3$ does not imply $F_2 = F_3$. If F is infinitely divisible, this is no longer possible. Many results concerning factorization, as well as references, are given by Lévy [4], [5].

To avoid trivialities, we presume in what follows that all c.d.f.'s have at least two points of increase and consider two c.f.'s $\phi_1(t)$ and $\phi_2(t)$ as equivalent if for some real α,

$$\phi_1(t) = \exp\{i\alpha t\}\, \phi_2(t).$$

The starting point of this investigation is the following

Definition. A family $\mathcal{S}$ of c.d.f.'s will be said to be *decomposable* ($\mathcal{S}'$) if, for any element F of $\mathcal{S}$, the relationship $F = G_1 * G_2$ implies that G_1 and G_2 are members of the family $\mathcal{S}'$. In particular, if $\mathcal{S} = \mathcal{S}'$, the family $\mathcal{S}$ will be called *factor-closed* (f.c.).[1]

Received 12/14/53.

[1] The class of all c.d.f.'s as well as the family of prime or indecomposable c.d.f.'s (i.e., the only "factors" of $\phi(t)$ are the trivial ones $\exp\{i\alpha t\}$ and $\phi(t)\exp\{-i\alpha t\}$) are trivially f.c.

Thus, Cramér's theorem [1], [2] on the factorization of the normal distribution states that the normal family is f.c. A corresponding result of Raikov [3] avers that the Poisson family is f.c.

For later usage, $P(z)$ is defined to be a *quasi-polynomial* if, for real d_j and r_j and integral $m \geqq 2$, $P(z) = \sum_{j=1}^{m} r_j z^{d_j}$. If, in addition, $m = 2$, $P(z)$ will be termed a *binomial* quasi-polynomial.

3. The general binomial family. We define a sequence of independent r.v.'s $\{X_j\}$, for $j = 1, 2, \cdots, n \geqq 2$, by

$$\Pr\{X_j = a_j\} = p_j, \qquad \Pr\{X_j = b_j\} = q_j = 1 - p_j,$$

where $a_j > b_j$ are real and $0 < p_j < 1$ for $j = 1, 2, \cdots, n$. Let

$$c_j = a_j - b_j > 0, \qquad V' = \sum_{j=1}^{n} X_j.$$

The c.f. of V' is

$$\phi_{V'}(t) = \prod_{j=1}^{n} (p_j e^{ita_j} + q_j e^{itb_j}) = \exp\left\{it \sum_{j=1}^{n} b_j\right\} \prod_{j=1}^{n} (p_j e^{itc_j} + q_j).$$

It suffices to consider the equivalent c.f.

$$\phi_V(t) = \prod_{j=1}^{n} (p_j e^{itc_j} + q_j) = A \prod_{j=1}^{n} (e^{itc_j} + \bar{q}_j)$$

where A is a constant and $\bar{q}_j = q_j p_j^{-1} > 0$. As $\phi_V(t)$ depends on the parameters a_j, b_j, p_j, and n, it represents a family of c.f.'s and there exists the corresponding family of c.d.f.'s, say $\mathfrak{J}$, whose explicit form is not required here. This family will be dubbed the *general binomial* family since it constitutes an obvious generalization of the classical binomial distributions connected with coin tossing, etc. It will be shown to be f.c. under certain conditions.

As the c.d.f. of X_j and hence of V is a step function with a finite number of jumps, the same must be true of any factor of the c.d.f. of V. We may therefore confine our attention (in looking for factors of $\phi_V(t)$) to c.f.'s of the form $\phi(t) = \sum_{j=1}^{m} r_j \exp\{itd_j\}$, where r_j is positive, d_j is real, and m is a positive integer $\geqq 2$. That is, we need only consider c.f.'s which are quasi-polynomials in $z = e^{it}$ with positive coefficients.

Lemma. *If a polynomial with nonnegative coefficients admits a factorization into quasi-polynomials with nonnegative coefficients, it admits a factorization into (ordinary) polynomials having the same coefficients.*

Proof. Let $P_0(z) = \prod_1^r P_i(z)$, where $P_0(z)$ is an ordinary polynomial with nonnegative coefficients and $P_i(z)$ is a quasi-polynomial for $i = 1, 2, \cdots, r$. Also, let m_i be the smallest exponent of $P_i(z)$ for $i = 0, 1, \cdots, r$. Since $\sum_1^r m_i$ is a nonnegative integer m_0, we have immediately $P_0'(z) = \prod_1^r P_i'(z)$, where $P_i'(z)$ is a quasi-polynomial with $m_i' = 0$ for $i = 0, 1, \cdots, r$. As any exponent appearing on the right side of the above equation must also appear on the left side, the $P_i'(z)$ must be ordinary polynomials. Q.E.D.

The distinguishing characteristic of the family $\mathfrak{J}$ is that $\phi_V(t)$ may be repre-

sented, by substituting $z = e^{it}$, as a product of binomial quasi-polynomials. If $\phi_V(t)$ may also be expressed as a product of quasi-polynomials which are not all reducible to binomial quasi-polynomials, it will be established that $\mathfrak{Z}$ is not, in general f.c. Consider the identity

$$(3) \qquad \begin{aligned}(z + 3)(z + 4)(z^3 + 8) &= (z + 2)(z^4 + 5z^3 + 2z^2 + 4z + 48)\\ &= (z + 2)\cdot P_4(z).\end{aligned}$$

Now, although $P_4(z)$ is in general reducible to $(z + 3)(z + 4)(z^2 - 2z + 4)$, it is irreducible into ordinary polynomials having nonnegative coefficients. By the lemma it is also irreducible to quasi-polynomials having nonnegative coefficients. If each parenthetic factor in (3) is divided by the sum of its coefficients and z is replaced by e^{it}, the expression on the left side is the c.f. of a member of $\mathfrak{Z}$, while that in the middle is a product of two c.f.'s, the second of which is not a member of $\mathfrak{Z}$.

On the other hand if $\mathfrak{Z}$ is suitably restricted, it is f.c. Let $\mathfrak{Z}_{c,2c}$ denote the subfamily of $\mathfrak{Z}$ with $c_j = c$ or $2c$ for $j = 1, 2, \cdots, n$. We have then

THEOREM 1. *The family $\mathfrak{Z}_{c,2c}$ is* f.c. *for any (positive) c.*

PROOF: If $c_j = 1$ or 2 for $j = 1, 2, \cdots, n$, then

$$\psi(z) = \phi_V(t) = A \prod_{j=1}^{n} (z^{c_j} + \bar{q}_j)$$

is the canonical decomposition of $\psi(z)$ into linear and quadratic factors. As $\bar{q}_j > 0$, it is clear that no matter how $\psi(z)$ is factored into ordinary polynomials, these must always be reducible to products of binomial factors with positive coefficients. With the lemma, this proves the theorem for the case $c = 1$. For arbitrary (positive) c, the transformation $y = z^c$ returns one to the case just examined.

COROLLARY 1. *Let $\mathfrak{Z}^p_{c,2c}$ denote the subfamily of $\mathfrak{Z}_{c,2c}$ wherein $p_j = p$ for $j = 1, 2, \cdots, n$. Then $\mathfrak{Z}^p_{c,2c}$ is* f.c.

PROOF: By Theorem 1, $\mathfrak{Z}^p_{c,2c}$ is decomposable $(\mathfrak{Z}_{c,2c})$. That $\mathfrak{Z}^p_{c,2c}$ is also f.c. follows directly from the fact that (for $c = 1$) all the roots of $\psi(z)$ must be equal to $-\bar{q}$ or $\pm i(\bar{q})^{1/2}$.

From Corollary 1, it follows that the only factors of the classical binomial (Bernoulli) distributions are themselves binomial distributions. It suffices here to choose $a_j \equiv 1$, $b_j \equiv 0$, and $p_j \equiv p$.

One might also define $\mathfrak{Z}^p$ as that subfamily of $\mathfrak{Z}$ for which $p_j = p$ for $j = 1, 2, \cdots, n$. However, it is simple to show via a counter-example that $\mathfrak{Z}^p$ is not f.c.

In generalization of the preceding, we define for any integral $k \geqq 2$ the general k-nomial family of distributions, say, U_k, as follows: Let $\{X_j\}$ be a sequence of independent random variables with

$$\Pr\{X_j = a_{ji}\} = p_{ji}, \qquad i = 1, 2, \cdots, k,$$

$$0 < p_{ji} < 1, \qquad \sum_{i=1}^{k} p_{ji} = 1, \qquad \text{all } j = 1, 2, \cdots, n.$$

There is no loss of generality in supposing $a_{j1} > a_{j2} > \cdots > a_{jk}$ for all j. Then $V' = \sum_1^n X_j$ will be a c.v. having the "general k-nomial distribution." We consider the case $k = 3$.

THEOREM 2. *If a_{ji} forms, for each j, an arithmetic progression whose common difference is independent of j, and if $p_{j2}^2 < 4p_{j1}p_{j3}$ for all j, then U_3 is* f.c.

PROOF. Let $b_{ji} = a_{ji} - a_{j3} > 0$ for $i = 1$ or 2. Then if $b_{j2} = b$, by hypothesis $b_{j1} = 2b$. As earlier, it suffices to consider

$$\phi_V(t) = \prod_{j=1}^n (p_{j1}e^{it2b} + p_{j2}e^{itb} + p_{j3}).$$

But this is the canonical decomposition of a polynomial in $W = e^{itb}$. In view of the positivity of the coefficients, and the lemma, U_3 is necessarily f.c. The conditions $p_{j2}^2 < 4p_{j1}p_{j3}$ preclude trivial decompositions into binomial distributions.

The factor-closedness of U_3 cannot be extended even to the case where the a_{ji} are in arithmetic progression but the difference depends on j. It suffices to note the counter-example

$$(z^6 + 30z^3 + {}^{6859}\!/_{27})\,(z^2 + 2z + 6)(z^2 + 3z + 6)$$
$$\equiv (z^2 + 5z + {}^{19}\!/_3)(z^4 + {}^{25}\!/_3 z^2 + {}^2\!/_3 z + 38)(z^4 + {}^{10}\!/_3 z^2 + z + 38).$$

4. The multinomial distribution. The factorization problem, as well as (1) and (2), extend readily to the m-dimensional case, that is, to m random variables or to a single vector random variable with m components. Where X, $F(x)$, and $\phi(t)$ were written previously, we need only substitute $(X_1, \cdots, X_m)$, $F(x_1, \cdots, x_m)$, and $\phi(t_1, \cdots, t_m)$. Cramér has shown [2] that the family of multivariate normal distributions is f.c.

We consider the classical multinomial distribution with n independent repetitions of an experiment whose m mutually exclusive and exhaustive outcomes $A_1, \cdots A_m$ have occurrence probabilities $p_1, \cdots p_m$. If X_j is the r.v. denoting the number of occurrences of A_j in the n trials, then

$$\Pr\{(X_1 = x_1), \quad \cdots, \quad (X_m = x_m)\} = \left(n! \Big/ \prod_{i=1}^m x_i!\right) p_1^{x_1}p_2^{x_2} \cdots p_m^{x_m},$$

where $\sum_1^m x_i = n$ and $\sum_1^m p_i = 1$. Here

$$\phi(t_1, t_2, \cdots, t_m) = [p_1e^{it_1} + p_2e^{it_2} + \cdots p_me^{it_m}]^n.$$

Let $z_j = e^{it_j}$ and $\psi(z_1, \cdots, z_m) = \phi(t_1, \cdots, t_m)$. As before, the only possible factors of ψ are of the form

$$\sum_{x_1} \cdots \sum_{x_m} q_{x_1,x_2,\cdots,x_m} z_1^{x_1}z_2^{x_2} \cdots z_m^{x_m} = \psi_1(z_1, z_2, \cdots, z_m). \tag{4}$$

Again there is no loss of generality in supposing the x_i to be nonnegative integers, that is, that ψ_1 is a polynomial rather than a quasi-polynomial. We now prove

THEOREM 3: *The family of (classical) multinomial distributions is* f.c.

PROOF: Analogous to (2), we have, where ψ_i is of the form (4) for $i = 1$ or 2,

$$(p_1z_1 + p_2z_2 + \cdots + p_mz_m)^n = \psi_1 \cdot \psi_2$$

Since the irreducible factor $(p_1z_1 + \cdots + p_mz_m)$ is an n-fold factor of $\psi_1 \cdot \psi_2$, it must be an n_1-fold factor of ψ_1 and an n_2-fold factor of ψ_2, with $n_1 + n_2 = n$, that is,

$$\psi_i = \left(\sum_{i=1}^{m} p_i z_i\right)^{n_i} Q_i(z_1, \cdots, z_m), \qquad i = 1, 2.$$

Clearly, Q_i = constant = 1, since $\psi(1, 1, \cdots, 1) = \phi(0, \cdots, 0) = 1$. Finally, $0 < n_j < n$ if degenerate c.f.'s and c.d.f's are precluded, as earlier.

Slight generalizations of Theorem 3 are possible. The writer has proved that the family of (correlated) multivariate Poisson distributions is f.c., but this will not be given here.

5. Infinitely divisible families. Returning to the unidimensional case, $F(x)$ is called *infinitely divisible* (i.d.) if $[\phi(t)]^{1/n}$ is a c.f. for every positive integer n. Khintchine's form [6] of Lévy's formula [4] gives as a necessary and sufficient condition for $F(x)$ to be i.d. that

$$\log \phi(t) = i\gamma t + \int_{-\infty}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1 + u^2}\right)\left(\frac{1 + u^2}{u^2}\right) dG(u) \tag{5}$$

where γ is real and $G(u)$ is bounded, monotonic nondecreasing and can be normalized so that $G(u^+) = G(u)$, $G(-\infty) = 0$, and $G(+\infty) = B$. Furthermore, the normalized representation is unique.

If $G(u)$ is a step function with only a single jump point, that is

$$\text{(a)} \quad G(u) = \begin{cases} 0, & u < 0, \\ \sigma^2, & u \geqq 0, \end{cases} \qquad \text{or} \qquad \text{(b)} \quad G(u) = \begin{cases} 0, & u < c \neq 0, \\ \frac{1}{2}\sigma^2, & u \geqq c, \end{cases}$$

then (a) yields the normal family of distributions while those in (b) are closely related to the Poisson family. If $c = 1$ and $\gamma = \frac{1}{2}\sigma^2$, then (b) is the Poisson family.

Suppose that $G(u)$ has n discontinuities $a_1 < a_2 < \cdots < a_n$ with saltuses $b_1, b_2, \cdots, b_n$. Then $G(u) = G_n(u) = G_n(u; a_1 \cdots a_n; b_1 \cdots b_n)$ has a corresponding i.d. c.f. $\phi_n(t; a_1 \cdots a_n; b_1 \cdots b_n)$ and c.d.f. $F_n(x; a_1 \cdots a_n; b_1 \cdots b_n)$.

For any fixed $n = 1, 2, \cdots$, consider the family

$$\mathfrak{F}_n = \{F_n(x; a_1 \cdots a_n; b_1 \cdots b_n)\}.$$

If $G_n(u)$ is a step-function, denote the corresponding family of c.d.f.'s by $\mathfrak{F}'_n$.

THEOREM 4. *For any fixed $n = 2, 3, \cdots$, the* i.d. *families $\mathfrak{F}_n$ and $\mathfrak{F}'_n$ are not* f.c.

PROOF. Let $b = \sum_{i=1}^{n-1} b_i$. Define

$$G_n^+(u) = \begin{cases} \dfrac{G_n(u) - b}{B - b}, & u \geqq a_n; \\ 0, & u < a_n; \end{cases} \qquad G_n^-(u) = \begin{cases} \dfrac{G_n(u)}{b}, & u < a_n, \\ 1, & u \geqq a_n. \end{cases}$$

Further, let $\phi_n^+(t)$ be given by (5) with $G_n^+(u)$ replacing $G(u)$, and define $\phi_n^-(t)$ analogously. Clearly, $G_n(u) = (B - b)G_n^+(u) + bG_n^-(u)$, whence

$$\phi_n(t; a_1 \cdots a_n; b_1 \cdots b_n) = \phi_n^+(t) \cdot \phi_n^-(t).$$

Since $G_n^+(u)$ has only one saltus, the c.d.f. corresponding to $\phi_n^+(t)$, say $F_n^+(x)$ belongs to either $\mathfrak{F}_1$ or $\mathfrak{F}_1'$. Thus, $\mathfrak{F}_n$ or $\mathfrak{F}_n'$ is not f.c. for $n \geqq 2$. In particular, if G_n is a step function, $F_n^+(x)$ is a normal or (almost) Poisson c.d.f.

REFERENCES

[1] H. Cramér, "Über eine Eigenschaft der normalen verteilungsfunktion," *Math. Z.*, Vol. 41 (1936), pp. 405–414.

[2] H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics, Cambridge, 1937.

[3] D. Raikov, "On the composition of Poisson laws," *C.R. Acad. Sci.* URSS, Vol. 14 (1937), pp. 8–11.

[4] P. Lévy, *Theorie de l'Addition des Variables Aléatoires*, Paris, 1937.

[5] P. Lévy, "L'arithmétique des lois des probabilités," *J. Math. Pures Appl.*, Vol. 17 (1928), pp. 17–39.

[6] A. Khintchine, "Contribution à l'arithmétique des lois de distribution," *Bull. Math. Univ. Moscou*, Vol. 1 (1937), pp. 6–17.

ON THE CONVOLUTION OF DISTRIBUTIONS

By Henry Teicher

Purdue University

1. Summary. A systematic approach to distributions having the reproductive property (see [1] p. 171) is attempted, and necessary and sufficient conditions are given. The case of distributions depending on k (> 1) parameters is considered; it need not be a straightforward generalization of the one-parameter case.

2. Additively closed families of distributions. Let $D = D(\lambda)$ be an Abelian semi-group under addition. In particular, denote by $D(I)$, $D(I+)$, $D(r+)$, $D(R+)$, and $D(R+, 0)$ the semi-groups of integers, positive integers, positive rationals, positive reals, and nonnegative reals, respectively. Let $D(r)$, $D(R)$, $D(I+, 0)$ and $D(R+, 0)$ be defined analogously. The abbreviations c.d.f. and c.f. will be used for cumulative distribution function and characteristic function, respectively.

Definition. A one-parameter family of c.d.f.'s $F(x; \lambda)$ with $\lambda \varepsilon D$, and D as above, will be said to be *additively closed* or to *belong to the class* C_1 if, for any two elements $F(x; \lambda_1)$ and $F(x; \lambda_2)$,

$$F(x; \lambda_1) * F(x; \lambda_2) \doteq F(x; \lambda_1 + \lambda_2). \tag{1}$$

Among the following results, Theorem 1 is known in one form or another but is required here for a unified presentation. Theorems 2 and 4 are new. Generally, the k-parameter case does not seem to have been considered previously.

Theorem 1. *If (i)* $\lambda \varepsilon D(I+)$ *or (ii)* $\lambda \varepsilon D(r+)$, *a necessary and sufficient condition that a family of* c.d.f.'s $F(x; \lambda)$ *be additively closed, that is, that* $F(x; \lambda) \varepsilon C_1$, *is that the corresponding family of* c.f.'s is $\phi(t; \lambda) = [f(t)]^\lambda$, *where* $f(t)$ *is a* c.f. *not depending on* λ. *If (iii)* $\lambda \varepsilon D(R+)$, *and* $\phi(t; \lambda)$ *is continuous in* λ, *the same condition is again necessary and sufficient. In cases (ii) and (iii),* $f(t)$ *is the* c.f. *of an infinitely divisible distribution.*

Proof. The proof of sufficiency is trivial for the ensuing theorems. The three alternatives for λ are considered in turn.

(i), $\lambda \varepsilon D(I+)$. Let $f(t) = \phi(t; 1)$. Translating and iterating (1), we have, for any positive p,

$$\phi(t; p) = \phi(t; 1)\phi(t; 1) \cdots \phi(t; 1) = [f(t)]^p.$$

(ii), $\lambda \varepsilon D(r+)$. We have, from (1), $f(t) = \phi(t; 1) = [\phi(t; 1/p)]^p$. That is, the pth root of $f(t)$ is a c.f. for every positive integral p, whence $f(t)$ is the c.f. of an infinitely divisible (i.d.) distribution and hence never zero. (By the pth root is meant that branch for which $f^{1/p}(0) = 1$, which is unambiguous since

Received 12/14/53.

$f(t) \neq 0$ for real t.) Again applying (1) we see that, for any positive integers p and q,

$$\phi(t; q/p) = [\phi(t; 1/p)]^q = [f(t)]^{q/p}.$$

(iii), $\lambda \varepsilon D(R+)$. Since $\phi(t; \lambda)$ is continuous in λ, it follows from (ii), by taking a sequence of positive rational numbers approaching any real nonnegative λ, that $\phi(t; \lambda) = [f(t)]^{\lambda}$.

If $\lambda \varepsilon D(R+)$ and the continuity assumption is removed, Theorem 1 is in general untrue. For example, let $F(x; \lambda)$ be a family of normal distributions with variance λ and mean $g(\lambda)$, where $g(\lambda)$ is a discontinuous solution of Cauchy's functional equation $g(x) + g(y) = g(x + y)$. Then

$$\phi(t; \lambda) = \exp\{itg(\lambda) - \tfrac{1}{2}\lambda t^2\}$$

is not of the form $[f(t)]^{\lambda}$ although $F(x; \lambda) \varepsilon C_1$.

THEOREM 2. *If $\phi(t; \lambda)$ for $\lambda \varepsilon D(R+)$ is real-valued (for real t), a* NSC *that $F(x; \lambda) \varepsilon C_1$ is that $\phi(t; \lambda) = [f(t)]^{\lambda}$.*

PROOF. The set of zeros of $\phi(t; \lambda)$ is independent of λ. For if $\phi(t_0; \lambda_1) = 0$ and $\lambda_2 > \lambda_1$, then

$$\phi(t_0; \lambda_2) = \phi(t_0; \lambda_2 - \lambda_1)\phi(t_0; \lambda_1) = 0. \tag{2}$$

If $\lambda_3 < \lambda_1$ and n is an integer, $[\phi(t_0; \lambda_1/n)]^n = \phi(t_0; \lambda_1) = 0$ whence $\phi(t_0; \lambda_1/n) = 0$ for every positive integer n. But for sufficiently large n, we have $\lambda_3 > \lambda_1/n$. Applying (2), we deduce $\phi(t_0; \lambda_3) = 0$.

For $\lambda = r$, a rational number, we have from Theorem 1 that $\phi(t; r) = [f(t)]^r$ with $f(t)$ never zero. It follows from the above that $\phi(t; \lambda)$ is never zero. Consequently, the properties of c.f.'s that $\phi(0; \lambda) = 1$ and that $\phi(t; \lambda)$ is continuous in t for every λ, show that $\phi(t; \lambda)$ is never negative.

Now $\psi(t; \lambda) = \log \phi(t; \lambda)$ is well defined, and, from the translated form of (1), satisfies Cauchy's functional equation. As

$$\phi(t; \lambda) = |\phi(t; \lambda)| \leqq 1,$$

$\psi(t, \lambda)$ is nonpositive whence the only solution is the continuous one $\psi(t; \lambda) = K_t\lambda$. Thus, for all real $\lambda > 0$,

$$\phi(t; \lambda) = \exp\{K_t\lambda\} = [h(t)]^{\lambda}.$$

Taking $\lambda = 1$, we have $\phi(t; 1) = h(t) = f(t)$, which proves the theorem.

DEFINITION. Let λ_j be an element of the Abelian semi-group (additive) D_j for $j = 1, 2, \cdots, k$. A k-parameter family of c.d.f.'s will be said to be *additively closed* or to belong to the class C_k if for any two members $F(x; \lambda_1^{(1)}, \cdots, \lambda_k^{(1)})$ and $F(x; \lambda_1^{(2)}, \cdots, \lambda_k^{(2)})$,

$$F(x; \lambda_1^{(1)}, \cdots, \lambda_k^{(1)}) * F(x; \lambda_1^{(2)}, \cdots, \lambda_k^{(2)}) \equiv F(x; [\lambda_1^{(1)} + \lambda_1^{(2)}], \cdots, [\lambda_k^{(1)} + \lambda_k^{(2)}]). \tag{3}$$

There may be a set of dormant parameters which are unaffected by the convolution, but these may simply be ignored.

In generalization of Theorem 1, we have:

THEOREM 3. *Let $F(x; \lambda_1, \cdots, \lambda_k)$ be a k-parameter family of* c.d.f.'s *with $\lambda_j \varepsilon D_j$ where $D_j = D_j(I+,0)$, $D_j(r+,0)$ or $D_j(R+,0)$. Further, let $\phi(t; \lambda_1, \cdots, \lambda_k)$ be continuous in all λ_j for which the corresponding $D_j = D_j(R+,0)$. Then a* NSC *that $F \varepsilon C_k$ is that*

$$\phi(t; \lambda_1, \lambda_2, \cdots, \lambda_k) = \prod_{j=1}^{k} [f_j(t)]^{\lambda_j},$$

where each $f_j(t)$ is a c.f. *independent of all λ_j, and is* i.d. *providing the corresponding D_j is $D_j(r+,0)$ or $D_j(R+,0)$.*

PROOF. As in Theorem 1, $\phi(t; 0, \cdots, 0, \lambda_j, 0, \cdots, 0) = G_j(t; \lambda_j) = [f_j(t)]^{\lambda_j}$. Hence,

$$\phi(t; \lambda_1, \cdots, \lambda_k) = \prod_{j=1}^{k} G_j(t; \lambda_j) = \prod_{j=1}^{k} [f_j(t)]^{\lambda_j}.$$

The inclusion of the value zero in each domain D_j immediately implies that each $f_j(t)$ is itself a c.f. The question arises whether this is necessarily so if zero is deleted. Provided the product space $D_1 \times D_2 \times \cdots \times D_k$ is suitably altered, the answer is in the negative.

THEOREM 4. *Let $F(x; \lambda_1, \lambda_2)$ be a two-parameter family of* c.d.f.'s *where $\lambda_1 \varepsilon D(r+)$ and $\lambda_2 \varepsilon D(r)$, with $\lambda_1 \geqq |\lambda_2|$ defining the parameter space. A* NSC *that $F(x; \lambda_1, \lambda_2) \varepsilon C_2$ is that*

$$\phi(t; \lambda_1, \lambda_2) = \prod_{j=1}^{2} [f_j(t)]^{\lambda_j},$$

where $f_2(t)$ is not necessarily a c.f.

PROOF. Since for any positive integer n,

$$[\phi(t; 1/n, 1/n]^n = \phi(t; 1, 1) = r(t), \quad \text{(say)},$$

$r(t)$ is an i.d.c.f., and $\phi(t; p/n, p/n) = [r(t)]^{p/n}$ for any positive integers p and n. Similarly,

$$[\phi(t; 1/m, 0)]^m = \phi(t; 1, 0) = f_1(t), \quad \text{(say)},$$

where $f_1(t)$ is an i.d.c.f. Hence $\phi(t; \lambda_1, 0) = [f_1(t)]^{\lambda_1}$ for $\lambda_1 \varepsilon D(r+)$. Let $f_2(t) = r(t)/f_1(t)$. Then $f_2(t)$ is defined and nonzero for all real t.

Now if $\lambda_2 > 0$ and $\lambda_1 = \lambda_2$, we have

$$\phi(t; \lambda_1, \lambda_2) = \phi(t; \lambda_1, \lambda_1) = [r(t)]^{\lambda_1} = [f_1(t)]^{\lambda_1}[f_2(t)]^{\lambda_2}.$$

If $\lambda_2 > 0$ but $\lambda_1 \neq \lambda_2$,

$$\phi(t; \lambda_1, \lambda_2) = \phi(t; \lambda_1 - \lambda_2, 0)\phi(t; \lambda_2, \lambda_2) = \prod_{j=1}^{2} [f_j(t)]^{\lambda_j}.$$

Furthermore,

$$\phi(t; \lambda_1, \lambda_2)\phi(t; \lambda_1, -\lambda_2) = \phi(t; 2\lambda_1, 0).$$

Substituting in this last equation and solving, we find

$$\phi(t; \lambda_1, -\lambda_2) = [f_1(t)]^{\lambda_1}[f_2(t)]^{-\lambda_2},$$

completing the proof. It is clear from the definition that $f_2(t)$ need not be a c.f.

The following example illustrates Theorem 4. Define $\phi_j(t) = \exp\{\alpha_j(e^{it} - 1)\}$ with $\alpha_1 > 0$ and $\alpha_2 \geqq 0$, and rational for $j = 1$ or 2. Let $\lambda_1 = \alpha_1 + \alpha_2$ and $\lambda_2 = \alpha_1 - \alpha_2$, with

$$\phi(t; \lambda_1, \lambda_2) = \phi_1(t)\phi_2(-t) = [e^{(\cos t)-1}]^{\lambda_1}[e^{i\sin t}]^{\lambda_2}. \tag{4}$$

The parameter space is given by $\lambda_1 \varepsilon D(r+)$ and $\lambda_2 \varepsilon D(r)$, with $\lambda_1 \geqq |\lambda_2|$. Finally, $\exp\{i \sin t\}$ cannot be a c.f. as

$$\exp\{i \sin t\} = 1 + it + \tfrac{1}{2}i^2t^2 + o(t^2), \tag{5}$$

which would imply that the corresponding r.v. had unit mean and zero variance and hence (by the uniqueness theorem for c.f.'s) a c.f. equal to $\exp\{it\}$.

The proof of the following generalization of Theorem 4 is very similar and will not be given.

THEOREM 5. *Let* $F(x; \lambda_1, \lambda_2, \cdots, \lambda_k)$ *be a k-parameter family of* c.d.f.'s, *where* $\lambda_1 \varepsilon D(r+)$ *and* $\lambda_j \varepsilon D(r+, 0)$, *with* $j \geqq 2$ *and* $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_k$ *defining the parameter space. A* NSC *that* $F(x; \lambda_1, \lambda_2, \cdots, \lambda_k) \varepsilon C_k$ *is*

$$\phi(t; \lambda_1, \lambda_2, \cdots, \lambda_k) = \prod_{j=1}^{k} [f_j(t)]^{\lambda_j}.$$

where $f_j(t)$ *is not necessarily a* c.f. *for* $j > 1$.

The last two theorems could be extended to real values of λ under suitable assumptions.

REFERENCES

[1] H. CRAMÉR, "Problems in probability theory," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 165–193.

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

NOTES

SEQUENTIAL PROCEDURES THAT CONTROL THE INDIVIDUAL PROBABILITIES OF COMING TO THE VARIOUS DECISIONS

By Lionel Weiss

University of Virginia

1. Summary. We consider cases where we have a finite number of decisions and a finite number of possible distributions, and we confine attention to procedures which have zero probability of continuing beyond the Nth observation, where N is a given positive integer. We find a class C of procedures such that given any procedure R, there is a member of C, say R', such that the probabilities of coming to the various decisions under the various distributions when using R' are at least as desirable as when using R, and such that we are at least as likely to take fewer than n observations under R' as under R, for any n. Various extensions are indicated.

2. Introduction. Since the discussion to follow leans very heavily on the results of a previous paper [1], we summarize briefly these results.

Let x be the generic point of a Euclidean space Z, and $F_1(x), \cdots, F_m(x)$ be m given cumulative probability distributions on Z. The statistician is confronted with an observation on the chance variable X which is distributed in Z according to an unknown one of $F_1, \cdots, F_m$. On the basis of this observation he has to make one of L decisions, say $d_1, \cdots, d_L$. Let s be a positive integer and $W_{ijk}(x)$, for $i = 1, \cdots, m$; $j = 1, \cdots, L$; $k = 1, \cdots, s$, be measurable functions of x such that $\int_Z |W_{ijk}(x)|\, dF_i(x) < \infty$.

A randomized decision function, often called "test" for short, and generically designated by $\eta(x)$, is defined as $\eta(x) = [\eta_1(x), \cdots, \eta_L(x)]$, where

(a) $\eta(x)$ is defined for all x;

(b) $0 \leqq \eta_j(x)$ for $j = 1, \cdots, L$;

(c) $\sum_1^L \eta_j(x) = 1$ identically in x;

(d) $\eta_j(x)$ is measurable for $j = 1, \cdots, L$.

Let

$$r_{ik} = \int_Z \left(\sum_{j=1}^{L} \eta_j(x) W_{ijk}(x) \right) dF_i(x),$$

$$r^s = (r_{ik}), \qquad i = 1, \cdots, m; \quad k = 1, \cdots, s$$

Thus to each $\eta(x)$ there corresponds the sth order *risk point* r^s. The test T with sth order risk point r^s will be said to be *uniformly better* (s) than the test T'

Received 12/18/53, revised 1/25/54.

with sth order risk point $r'^s = (r'_{ik})$ if $r_{ik} \leqq r'_{ik}$ for every i and k, with the inequality sign holding for at least one pair (i, k). A test T will be called *admissible* (s) if there exists no test uniformly better (s) than T. A class C_0 of tests will be called *complete* (s) if, for any test T' not in C_0, there exists a test T in C_0 which is uniformly better (s) than T'.

Any set $\xi = (\xi_{ik})$, for $i = 1, \cdots, m$, and $k = 1, \cdots, s$, of nonnegative numbers which add to unity (a convenient normalization) will be called an *a priori distribution* (s). A *Bayes' solution* (s) with respect to ξ is a test T_0 which minimizes $\sum_{i,k} \xi_{ik} r_{ik}(T)$ with respect to all tests T.

Let f_i be the density function of F_i with respect to a measure μ, with respect to which all F_i are absolutely continuous. There is always such a measure. To construct an sth order Bayes' solution with respect to $\xi = (\xi_{ik})$ one may proceed as follows: $\eta_j(x) = 0$ for all $j = 1, \cdots, L$ for which

$$\sum_{i=1}^{m} \sum_{k=1}^{s} \xi_{ik} f_i(x) W_{ijk}(x)$$

is not a minimum with respect to j; $\eta_j(x)$ is defined arbitrarily between 0 and 1, inclusive, for all other j, provided only that every component of the resulting $\eta(x)$ is measurable and the sum is always 1.

The only other result of the previous paper [1] needed for this one is the complete class theorem: *Every admissible* (s) *test is a Bayes' solution* (s) *with respect to some a priori distribution* (s). *Hence the class of Bayes' solutions* (s) *is complete* (s).

The present paper deals with the case where there is a finite number of terminal decisions, $d_1, \cdots, d_L$, one of which must be chosen, and a finite number of possible distributions for the vector chance variable $X = (X_1 \cdots X_N)$. We observe X sequentially, that is first we observe X_1, then decide whether to observe X_2 or choose a terminal decision on the basis of X_1 alone, etc. Each component X_i of X may itself be a vector chance variable, and $X_1, \cdots, X_N$ are not necessarily independent. For each distribution $F_j(x)$, we assume that the terminal decisions are divided into two mutually exclusive, exhaustive, and nonempty classes, one of "favorable" decisions (those we prefer to make when $F_j(x)$ is the true distribution) and the rest "unfavorable".

At various points in the discussion to follow, certain functions are well-defined except on sets of measure zero. The existence of these exceptional sets is of no consequence, and will not be specifically mentioned. As usual, $P(A \mid B)$ will denote the conditional probability of A given B.

3. A class of optimum decision procedures. Suppose that we are examining some specific decision procedure R. Once R tells us to stop sampling after observing X_n, R chooses a terminal decision in some way. We replace this way by a Bayes' solution (L) to the decision problem, where the space is the set of all points $(x_1 \cdots x_n)$. The possible distributions $\bar{F}_1, \cdots, \bar{F}_m$ are such that, over any Borel set S in the space of $(x_1 \cdots x_n)$, the integral $\int_S d\bar{F}_i$ is the

probability that sampling stops at a point of S, using R and F_i true, given that sampling stops with x_n. The decisions are $d_1, \cdots, d_L$. The function $W_{ijk}(x_1 \cdots x_n)$ is 0 if d_k is favorable relative to F_j and $j = k$, or if d_k is unfavorable and $j \neq k$; it is 1 otherwise, that is if d_k is favorable and $j \neq k$, or if d_k is unfavorable and $j = k$. The complete class theorem quoted above shows that this Bayes' solution (L) can be so chosen that the probabilities of all the unfavorable decisions are no greater than they were when using R, and the probabilities of all the favorable decisions are at least as great as they were when using R, under all distributions.

We note that we have not yet changed the stopping rule, which is still given by the originally specified procedure R. We also note that the a priori distributions (L) will in general be different for different n. Let this a priori distribution (L) as a function of n be denoted by ${}_n\xi = ({}_n\xi_{ik})$ for $i = 1, \cdots, m$ and $k = 1, \cdots, L$. We denote by

${}_nF_i(x_1 \cdots x_n)$ the marginal cumulative distribution function given by $F_i(x_1 \cdots x_N)$ for $(X_1 \cdots X_n)$;

${}_nf_i(x_1 \cdots x_n)$ the density function of ${}_nF_i(x_1 \cdots x_n)$ with respect to a measure μ, with respect to which all ${}_nF_i$ are absolutely continuous; and

$s(x_1 \cdots x_n)$ the probability under the stopping rule given by R that sampling will be stopped exactly after observing X_n when the first n observed values are $(x_1 \cdots x_n)$.

Then we have that the density function for $\bar{F}_i(x_1 \cdots x_n)$ is

$$\frac{s(x_1 \cdots x_n)\ {}_nf_i(x_1 \cdots x_n)}{\int s(x_1 \cdots x_n)\ {}_nf_i(x_1 \cdots x_n)\, d\mu}.$$

If the denominator of this expression is zero for some value of i, then we consider that in this particular problem of choosing a terminal decision, $\bar{F}_i$ is not one of the possible distributions, and we modify our class of possible decision procedures accordingly. Now we note that according to the explicit construction of a Bayes' solution (L) given in Section 2, the Bayes' solution (L) with respect to ${}_n\xi$ that we are using is identical with a Bayes' solution (L) to the problem with the same decisions and the same functions $W_{ijk}(x)$, but with distributions given by the density functions ${}_nf_1, \cdots, {}_nf_m$, (leaving out any corresponding to zero denominators in the fraction above) and with a priori distribution ${}_n\xi' = ({}_n\xi'_{ik})$, where

$${}_n\xi'_{ik} = \frac{C\ {}_n\xi_{ik}}{\int s(x_1 \cdots x_n)\ {}_nf_i(x_1 \cdots x_n)\, d\mu},$$

C being a normalizing constant, and it being understood that any fraction with a zero denominator is to be zero. Below we shall modify our stopping rule, but until further notice, whenever it is decided to stop sampling with X_n under any stopping rule, we make our terminal decision by means of the Bayes' solu-

tion (L) with respect to the a priori distribution ${}_n\xi'$, where the distributions are given by the densities ${}_nf_1, \cdots, {}_nf_m$, just described.

Now we construct a (possibly) new stopping rule. For a given n between 1 and $N - 1$ inclusive, to decide whether or not to stop after observing X_n, let us assume that we have already described a rule for stopping after observing X_{n+1}, and $X_{n+2}, \cdots$, and X_{N-1}, while we use the stopping rule given by R before we reach X_n. We look upon the problem of deciding whether to continue sampling after observing $(x_1 \cdots x_n)$ as a decision problem with two possible decisions: D_1, stop sampling; and D_2, continue sampling.

We apply the complete class theorem quoted above to this case with the decisions D_1 and D_2. The distributions, given by the density functions ${}_n\hat{f}_1(x_1 \cdots x_n), \cdots, {}_n\hat{f}_m(x_1 \cdots x_n)$, are defined as follows. Let $t(x_1 \cdots x_n)$ denote the probability that sampling will not be stopped before observing X_n when the procedure R is used and the first n observed values are $(x_1 \cdots x_n)$. Then set

$${}_n\hat{f}_i(x_1 \cdots x_n) = \frac{t(x_1 \cdots x_n)\ {}_nf_i(x_1 \cdots x_n)}{\int t(x_1 \cdots x_n)\ {}_nf_i(x_1 \cdots x_n)\, d\mu}.$$

The functions $W_{ijk} = W_{ijk}(x_1 \cdots x_n)$, for $i = 1, \cdots, m$; $j = 1, 2$; and $k = 1, \cdots, N - n + L$, are defined as follows:
for $k = 1$,

$$W_{i11} = 0, \qquad W_{i21} = 1, \qquad \text{for all } i, \text{ all } (x_1 \cdots x_n);$$

for $k = 2, \cdots, N - n$,

$$W_{i1k} = 0 \qquad \text{for all } i, \text{ all } (x_1 \cdots x_n);$$

$$W_{i2k} = P[\text{not stopping before } X_{n+k} \text{ under } F_i \mid (x_1 \cdots x_n)];$$

for $k = N - n + 1, \cdots, N - n + L$,

$$W_{ijk} = P[\text{after } D_j \text{ is made of } \{{}^{\text{not}}\} \text{ making } d_{k-N+n} \text{ under } F_i \mid (x_1 \cdots x_n)],$$

$$\text{if } d_{k-N+n} \text{ is } \{{}_{\text{un}}\} \text{ favorable relative to } F_i.$$

All the conditional probabilities introduced here in defining $W_{ijk}(x)$ are well-defined, since we assumed that we have already decided how to proceed once we decide either to stop after observing X_n or to continue sampling after X_n. By the complete class theorem quoted above, there is an a priori distribution $(N - n + L)$ for this two-decision problem, with a Bayes' solution $(N - n + L)$ no worse than the solution given by the procedure R with respect to the probabilities of making favorable or unfavorable terminal decisions and with respect to the distribution of the sample size required to come to a decision, under any of the possible distributions.

We apply this construction first to the problem of deciding whether to stop after observing X_{N-1}. Let ${}_{N-1}\xi$ be the a priori distribution $(L + 1)$ to be used. We define an a priori distribution ${}_{N-1}\bar{\xi}'$ and a Bayes' solution for it in terms of

${}_{N-1}\bar{\xi}$, just as we defined ${}_{N-1}\xi'$ in terms of ${}_{N-1}\xi$ above, using $t(x_1 \cdots x_{N-1})$ instead of $s(x_1 \cdots x_{N-1})$. Then we agree that no matter how the stopping rule may be modified before reaching X_{N-1}, we will decide whether to continue sampling after reaching X_{N-1} by using this Bayes' solution $(L + 1)$ with respect to ${}_{N-1}\bar{\xi}'$. Here the decisions are D_1 and D_2, the distributions are given by the densities ${}_{N-1}f_1(x_1 \cdots x_{N-1}), \cdots, {}_{N-1}f_m(x_1 \cdots x_{N-1})$, and $W_{ijk}(x)$ are as given above. Now we apply the construction to the problem of deciding whether to stop sampling after observing X_{N-2}, carrying out similar steps and getting an a priori distribution ${}_{N-2}\bar{\xi}'$, etc., down to deciding whether to stop sampling after observing X_1, getting an a priori distribution ${}_1\bar{\xi}'$.

The net result of all this is to give us a decision procedure R' which uses higher order Bayes' solutions to decide whether to continue sampling at each stage, and also to choose a terminal decision after it is decided to stop sampling. This R' is such that for any distribution F_i, the probability of making any given favorable decision is at least as high using R' as using R, and the probability of making any particular unfavorable decision is no higher under R' than under R. Also, under any F_i, the cumulative distribution function of the sample size required to come to a terminal decision when using R' is never below the corresponding function when using R.

Now let C be the class of decision procedures we get by varying all the various a priori distributions used in defining R' (both those used in setting the stopping rule and those used in choosing terminal decisions) in all possible ways, modifying the Bayes' solutions accordingly. Since R was arbitrary, we have the theorem that for any decision procedure R there is a member of C, say R', enjoying the same advantages over R as were described in the preceding paragraph.

4. Extensions. The individual probabilities of making terminal decision d_j under distribution F_i after no more than n observations can be controlled for all i, j, and n in a manner similar to that described above, by defining the proper set of functions $W_{ijk}(x)$ at each stage. How this would be done is clear, and is not elaborated here.

Obvious analogues to the above results hold for cases, such as two-sample problems, where the size of the second sample is a bounded function of the observations in the first sample. Choice of the size of the second sample is treated as a decision problem, just as whether to continue sampling was treated as a decision problem above.

The results of [2] can be extended to the present case, so that when we have atomless distributions our class C above need not contain procedures employing randomization.

Under certain conditions, a case where there is an infinite number of distributions and/or decisions and/or possible sample sizes can be approximated arbitrarily closely by a case where these three numbers are finite [3].

Finally, cases in which it is desired to control not the whole distribution of

the sample size required to come to a terminal decision, but only certain aspects of it (for example, its expected value), can be handled as above, using the proper $W_{ijk}(x)$ at each stage.

REFERENCES

[1] LIONEL WEISS, "A higher order complete class theorem," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 677-680.
[2] A. DVORETZKY, A. WALD, AND J. WOLFOWITZ, "Elimination of randomization in certain statistical decision procedures and zero-sum two-person games," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 1-21.
[3] J. WOLFOWITZ, "On ϵ-complete classes of decision functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 461-465.

ON A CHARACTERISATION OF THE GAMMA DISTRIBUTION

BY R. G. LAHA

Indian Statistical Institute

An intrinsic property of the gamma distribution, as proved by Pitman [1], is that if $X_1, X_2 \cdots, X_n$ are n identically distributed independent gamma variates with the distribution function

$$dF(X) = \frac{1}{\Gamma(p)} e^{-X} X^{p-1} \, dX \qquad (0 \leqq X \leqq \infty)$$

then the sum $X_1 + X_2 + \cdots + X_n$ is distributed independently of any function $g(X_1, X_2, \cdots, X_n)$ satisfying $g(X_1, X_2, \cdots, X_n) = g(\lambda X_1, \lambda X_2, \cdots, \lambda X_n)$ for any nonzero real λ. That is, $g(X_1, X_2, \cdots, X_n)$ should be a function independent of scale. In the present paper the converse theorem is proved for a particular class of g-function.

THEOREM. *Let $X_1, X_2, \cdots, X_n$ be n identically distributed independent random variables with a finite second moment. If the conditional expectation of the ratio of two quadratic forms $(\sum a_{ij}X_iX_j)/(\sum X_i)^2$, (where the elements of the matrix (a_{ij}) satisfy the relation $\sum a_{ii} \neq \sum a_{ij}/n$) for fixed sum $X_1 + X_2 + \cdots + X_n$ be equal to its unconditional expectation, then each X follows the gamma distribution.*

For a matrix $A = (a_{ij})$ where the relation $\sum a_{ii} = \sum a_{ij}/n$ holds, the method suggested does not lead to any solution of the problem. It is also interesting to note in this connection that the stronger assumption of stochastic independence of the sum $X_1 + X_2 + \cdots + X_n$ and $g(X_1, X_2, \cdots X_n)$ is not necessary for this particular class of g-function.

The following lemma is required for the proof of the Theorem.

LEMMA. *If u and v are two random variables such that for fixed v, the conditional expectation of $u/f(v)$, where $f(v)$ is a function of v, is equal to its unconditional expectation (provided it exists), then*

$$E(ue^{itv}) = E\{u/f(v)\} \cdot E\{f(v)e^{itv}\}.$$

Received 9/14/53, revised 5/28/54.

The proof of this lemma is very simple. If x and y are two variates such that the conditional expectation of y for fixed x is equal to its unconditional expectation, then $E_x(y) = E(y)$. Multiplying both sides by $\varphi(x)\, e^{itx}$ and taking expectation with respect to x, we get very easily

$$E\{y\, \varphi(x) e^{itx}\} = E(y) \cdot E\{\varphi(x)\, e^{itx}\}.$$

Putting $y = u/f(v)$, $\varphi(x) = f(v)$, $x = v$, this becomes

$$E\{(u/f(v)) f(v) e^{itv}\} = E\{u/f(v)\} \cdot E\{f(v) e^{itv}\}.$$

To prove the lemma, this may be written as

$$E\{u e^{itv}\} = E\{u/f(v)\} \cdot E\{f(v) e^{itv}\}.$$

PROOF OF THEOREM. Using this lemma with $u = \sum a_{ij} X_i X_j$, $v = \sum X_i$, and $f(v) = (\sum X_i)^2$,

$$\begin{aligned} &E\{(\textstyle\sum a_{ij} X_i X_j) e^{it(X_1 + X_2 + \cdots + X_n)}\} \\ &\qquad = E\{(\textstyle\sum a_{ij} X_i X_j)/(\sum X_i)^2\} \cdot E\{(\sum X_i)^2 \cdot e^{it(X_1 + \cdots + X_n)}\}. \end{aligned} \tag{1}$$

Let $\varphi(t) = E(e^{itX})$ represent the characteristic function of the distribution of X. After some algebraic simplifications, (1) will reduce to

$$\begin{aligned} &(\textstyle\sum a_{ii}) \cdot \dfrac{d^2\varphi}{dt^2} \cdot \varphi^{n-1} + (\sum\limits_{i \neq j} a_{ij}) \cdot \left(\dfrac{d\varphi}{dt}\right)^2 \cdot \varphi^{n-2} \\ &\qquad = K\left\{n \cdot \dfrac{d^2\varphi}{dt^2} \cdot \varphi^{n-1} + n(n-1) \cdot \left(\dfrac{d\varphi}{dt}\right)^2 \cdot \varphi^{n-2}\right\}, \end{aligned} \tag{2}$$

where $K = E\{(\sum a_{ij} X_i X_j)/(\sum X_i)^2\}$. Then, we have

$$\begin{aligned} &(\textstyle\sum a_{ii}) \cdot \left(\dfrac{d^2\varphi}{dt^2} \Big/ \varphi\right) + (\sum\limits_{i \neq j} a_{ij}) \cdot \left(\dfrac{d\varphi}{dt} \Big/ \varphi\right)^2 \\ &\qquad = K\left\{n \cdot \left(\dfrac{d^2\varphi}{dt^2} \Big/ \varphi\right) + n(n-1) \cdot \left(\dfrac{d\varphi}{dt} \Big/ \varphi\right)^2\right\}. \end{aligned} \tag{3}$$

Writing $\psi(t) = \ln \varphi(t)$, we have

$$\frac{d\psi}{dt} = \frac{d\varphi}{dt} \Big/ \varphi, \qquad \frac{d^2\psi}{dt^2} = \frac{d^2\varphi}{dt^2} \Big/ \varphi - \left(\frac{d\varphi}{dt} \Big/ \varphi\right)^2.$$

Substituting these in (3), we obtain the following differential equation for $\psi(t)$,

$$A \cdot \frac{d^2\psi}{dt^2} + B \cdot \left(\frac{d\psi}{dt}\right)^2 = 0, \qquad \begin{cases} A = \sum a_{ii} - nK, \\ B = \sum a_{ij} - n^2 K, \end{cases} \tag{4}$$

together with the initial conditions

$$\left.\frac{d\psi}{dt}\right|_{t=0} = im, \qquad \left.\frac{d^2\psi}{dt^2}\right|_{t=0} = -\sigma^2.$$

Here m and σ^2 are respectively the mean and variance of the distribution of X. In the solution of this differential equation (4), three cases must be distinguished.

I. $A \neq 0, \quad B \neq 0;$

II. $A \neq 0, \quad B = 0;$

III. $A = 0, \quad B \neq 0.$

For Case I, the differential equation may be written as

$$\frac{d^2\psi}{dt^2} = C \cdot \left(\frac{d\psi}{dt}\right)^2, \quad C = -\frac{B}{A} = \frac{\sigma^2}{m^2}, \tag{5}$$

using the initial condition in (4). Writing $\xi(t) = d\psi/dt$ equation (5) reduces to

$$\frac{d}{dt}\left(\frac{1}{\xi(t)}\right) = -\frac{\sigma^2}{m^2}. \tag{6}$$

Integrating this differential equation (6) with respect to t, using the initial condition $\xi(0) = im$, we get

$$\frac{1}{\xi(t)} = \frac{1}{im} - \frac{\sigma^2}{m^2}t, \quad \text{or} \quad \xi(t) = \frac{im}{1 - (\sigma^2/m)it}. \tag{7}$$

From (7), with the initial condition $\psi(0) = 0$, we get very easily

$$\psi(t) = -(m^2/\sigma^2)\log[1 - (\sigma^2/m)it], \quad \text{or} \quad \varphi(t) = [1 - (\sigma^2/m)it]^{-(m^2/\sigma^2)}. \tag{8}$$

By applying the inversion theorem, it can be very easily shown that the characteristic function $\varphi(t)$ in (8) leads uniquely to the gamma distribution with parameters $\alpha = m/\sigma^2$ and $\beta = m^2/\sigma^2$, the frequency function being given by

$$\begin{cases} [1/\Gamma(\beta)]\alpha^\beta \cdot e^{-\alpha X}X^{\beta-1} & X > 0 \\ 0 & X \leqq 0 \end{cases} \quad m > 0; \qquad \begin{cases} 0 & X \geqq 0 \\ [1/\Gamma(\beta)](-\alpha)^\beta e^{-\alpha X}(-X)^{\beta-1} & X < 0 \end{cases} \quad m < 0. \tag{9}$$

For cases II and III, it follows from the conditions stated in the theorem that

$$E\left\{\frac{\sum a_{ij}X_iX_j}{(\sum X_i)^2}\right\} = \frac{\sigma^2\sum a_{ii} + m^2\sum a_{ij}}{n\sigma^2 + n^2m^2}. \tag{10}$$

Thus the condition $B = 0$ yields the relation

$$\frac{\sum a_{ij}}{n^2} = K = E\left\{\frac{\sum a_{ij}X_iX_j}{(\sum X_i)^2}\right\} = \frac{\sigma^2\sum a_{ii} + m^2\sum a_{ij}}{n\sigma^2 + n^2m^2}. \tag{11}$$

On simplification, this reduces to $\sum a_{ii} = \sum a_{ij}/n$. Similarly, in Case III, the condition $A = 0$ obviously leads to the relation

$$\frac{\sum a_{ii}}{n} = K = \frac{\sigma^2\sum a_{ii} + m^2\sum a_{ij}}{n\sigma^2 + n^2m^2}. \tag{12}$$

On simplification, this also reduces to $\sum a_{ii} = \sum a_{ij}/n$, the same as obtained from the condition $B = 0$. Thus an important conclusion is reached that whenever the matrix $A = (a_{ij})$ is such that its elements satisfy the relation $\sum a_{ii} = \sum a_{ij}/n$ both the coefficients A and B of the differential equation (4) vanish simultaneously, thus leading to no solution of the problem.

Since cases II and III are excluded by our assumption $\sum a_{ii} \neq \sum a_{ij}/n$, the problem leads uniquely to the solution obtained in (9). Obviously when the matrix $A = (a_{ij})$ is either positive definite or negative definite, the relation $\sum a_{ii} \neq \sum a_{ij}/n$ is always satisfied. Thus the equality $\sum a_{ii} = \sum a_{ij}/n$ may hold only for some indefinite matrices.

COROLLARY. *Let $X_1, X_2, \cdots, X_n$ be identically distributed independent random variables with a finite second moment. If the ratio of the linear functions of random variables given by $(a_1X_1 + \cdots + a_nX_n)/(X_1 + \cdots + X_n)$ is distributed independently of the sum $X_1 + X_2 + \cdots + X_n$ then each X will follow a gamma distribution.*

PROOF. From the statement above, it follows that the conditional expectation of $(a_1X_1 + \cdots + a_nX_n)^2/(X_1 + \cdots + X_n)^2$ for the fixed sum $X_1 + \cdots + X_n$ is equal to its unconditional expectation. Here the elements of the matrix A are given by $a_{ij} = a_ia_j$ for $i,j = 1, 2, \cdots, n$ and they always satisfy the Schwartz's inequality $\sum a_i^2 > (\sum a_i)^2/n$, excluding the trivial case $\sum a_i^2 = (\sum a_i)^2/n$ which is possible when and only when all a_i's are equal, thus reducing the ratio of the linear functions to a constant. Hence the relation $\sum a_{ii} \neq \sum a_{ij}/n$ is always satisfied and the proof follows at once.

REFERENCE

[1] E. J. G. PITMAN, "The 'closest' estimates of statistical parameters," *Proc. Cambridge Philos. Soc.*, Vol. 33 (1937), pp. 212–222.

MATCHING IN PAIRED COMPARISONS

BY J. L. HODGES, JR. AND E. L. LEHMANN[1]

University of California, Berkeley

1. One of the simplest designs for testing the effect of a treatment is the method of paired comparisons: $2n$ subjects are divided into n pairs, and within each pair the treatment is assigned at random to one of the two subjects while the other is used as a control. This method has the reputation of being most effective if the subjects within each pair are as closely matched as possible. We shall show below that while this is true in the situations occurring most commonly in practice, it is not correct universally.

Received 11/2/53.

[1] This paper was prepared with the partial support of the Office of Naval Research.

We are interested in the power of the one-sided sign test for testing the hypothesis H of no effect against the simple alternative K that the treatment has a specified positive effect.

Consider now a possible pair of subjects and assume the usual model: the score of A, B is composed of a true value a, b, an error term U, V and, in case the treatment is applied and is effective, a treatment effect t. Then if X and Y are the scores of A and B, respectively, we have under H: $X = a + U$, $Y = b + V$, while under K the quantity t is added to the score of the treated subject. We assume that U and V are identically and independently distributed according to a continuous distribution F, and denote by G the distribution of $V - U$. Then if the treatment is applied to A or B, with probability $\frac{1}{2}$ each, the probability that the score of the treated subject exceeds that of the untreated one is $\frac{1}{2}$ under H, and

$$\tfrac{1}{2}[G(t + \Delta) + G(t - \Delta)]$$

under K, where $\Delta = b - a$. Without loss of generality, Δ may be taken as nonnegative.

If A and B are perfectly matched, then $\Delta = 0$ and the probability that the treated subject has the greater score becomes $G(t)$. Perfect matching can therefore be guaranteed to give the highest power against all alternatives if and only if

$$(1) \qquad \tfrac{1}{2}[G(t + \Delta) + G(t - \Delta)] \leqq G(t) \qquad \text{for all } t \geqq 0, \text{ all } \Delta.$$

This condition clearly implies that $G(u)$ is concave for $u \geqq 0$: that the converse is also true is at once obvious for $\Delta \leqq t$. For $t < \Delta \leqq 2t$, note that the values of G involved in (1) are unaltered if in the interval $[t - \Delta, \Delta - t]$ the function G is replaced by its chord. The resulting curve is concave to the right of $t - \Delta$ and (1) follows. Finally, for $\Delta > 2t$, we note that (1) is equivalent to

$$(2) \qquad G(\Delta + t) - G(\Delta - t) \leqq G(t) - G(-t) \qquad \text{for all } t \geqq 0,\ \Delta \geqq 0.$$

This time replace G by its chord in the interval $[-t, t]$, to establish (1).

Matters simplify if we assume that G has a density g. Then the convexity of G is equivalent to the requirement that the symmetrical function $g(u)$ be a decreasing for $u \geqq 0$, and hence unimodal (with mode 0). In summary, a necessary and sufficient condition for perfect matching to give always the greatest power is that the density g be unimodal.

It is clear that there are distributions F of the error U for which this condition holds. The normal case is an example, since then G is again a normal distribution. However, it is also easy to give examples for which the condition is not satisfied. Let F be uniformly distributed over the union of the intervals (0, 1) and (4, 5). Then $g(u) = 0$ for $1 < |u| < 3$ and is positive for $3 < |u| < 5$. In this extreme example the gain in power may be considerable. We have $G(1) = G(3) = \frac{3}{4}$ and $G(5) = 1$. With $t = 3$ the probability that the treated subject exceeds the untreated one is $\frac{3}{4}$ when $\Delta = 0$ and $\frac{7}{8}$ when $\Delta = 2$. If we use 10

pairs and consider the treatment as significant when the response of the treated subject is higher in eight or more pairs, the significance level is .055. The power against a treatment-effect $t = 3$ is then only .526 when identical subjects are paired but rises to .880 when the subjects in each pair have a response difference $\Delta = 2$. Thus, for certain error distributions and sizes of treatment effects, it is possible to improve the power of the test substantially by purposely mismatching.[2]

It appears that to use the possibility of improving the power (when it exists), one must know the distribution G. But if G were known, one could obtain a more powerful test based on the differences themselves, instead of just on the signs of differences. This is the very common difficulty, that the choice of an optimum design depends on knowledge which a priori was assumed unavailable. However, while values of nuisance parameters, form of distributions, etc., frequently are not sufficiently well known for the validity of the test to depend on this knowledge, one does have some idea about them, which may be utilized in the design of the experiment. The statistical procedure then will be valid, whether one's ideas are correct or not. Only the sensitivity of the experiment will be affected by the accuracy of these ideas.

In the next section we shall show that g is unimodal whenever F has a unimodal density, and this is the case in most applications. However, bimodal error distributions do occur, particularly when there is the possibility of "gross error." In such cases mismatching may increase the power of the test.

2. The purpose of this section is to prove that the difference of two independent observations on a unimodal random variable has also a unimodal distribution. We note that the same is not true of the sum, as has been pointed out by Chung [1], who gives a counter example. It is also easy to see that our condition is not a necessary one by considering

$$P(X = 1) = 1/5 \text{ and } P(X = 0) = P(X = 2) = 2/5.$$

DEFINITION. We say that a random variable X is unimodal with mode m

(a) in the discrete case, if the possible values of X are equally spaced numbers $m, m \pm \Delta, m \pm 2\Delta, \cdots$, and

$$\cdots \leqq P(X = m - 2\Delta) \leqq P(X = m - \Delta) \leqq P(X = m)$$
$$\geqq P(X = m + \Delta) \geqq P(X = m + 2\Delta) \geqq \cdots,$$

(b) or, in the continuous case, if X has a density function f which is increasing for $x < m$ and decreasing for $x > m$.

We shall need the following inequality.

THEOREM 1. *Let $(a_1, a_2, \cdots, a_n)$ be a sequence of real numbers satisfying*

$$0 \leqq a_1 \leqq \cdots \leqq a_m \geqq a_{m+1} \geqq \cdots \geqq a_n \geqq 0 \tag{3}$$

[2] It should, however, be pointed out that the corresponding possibility does not exist if one is interested in a point estimate of the treatment effect.

for some $1 \leqq m \leqq n$. *Let* $S_k = a_1a_{1+k} + a_2a_{2+k} + \cdots + a_{n-k}a_n$ *for* $k = 0, 1, \cdots, n-1$. *Then* $S_0 \geqq S_1 \geqq \cdots \geqq S_{n-1}$.

PROOF. Fix $k \geqq 0$ and prove $S_k \geqq S_{k+1}$ for $n \geqq k + 2$. For $n = k + 2$ our proposition becomes

$$a_1a_{k+1} + a_2a_{k+2} \geqq a_1a_{k+2},$$

which is easily verified: $a_1a_{k+1} \geqq a_1a_{k+2}$ unless $a_{k+1} < a_{k+2}$, in which case $a_1 \leqq a_2$ and $a_2a_{k+2} \geqq a_1a_{k+2}$. We induct on n. Let there be given any sequence $(a_1, \cdots, a_n)$ satisfying (3), with $n > k + 2$. We may assume $m > 1 + k$, since otherwise we have easily

$$S_k - S_{k+1} = a_1(a_{1+k} - a_{2+k}) + \cdots + a_{n-k-1}(a_{n-1} - a_n) + a_{n-k}a_n \geqq 0.$$

Since we also have

$$S_k - S_{k+1} = a_1a_{1+k} + (a_2 - a_1)a_{2+k} + \cdots + (a_{n-k} - a_{n-k-1})a_n,$$

the theorem is obvious if $m \geqq n - k$. We therefore now assume $1 + k < m < n - k$.

Let us consider the sequence $(a_1, \cdots, a_{m-1}, a_{m+1}, \cdots, a_n)$ obtained from the given sequence by dropping a_m, and let S' denote the sums of products for the new sequence. Note that the new sequence also satisfies (3). We have

$$\begin{aligned} S_k' &= (a_1a_{1+k} + \cdots + a_{m-1-k}a_{m-1}) + (a_{m-k}a_{m+1} + \cdots + a_{m-1}a_{m+k}) \\ &\quad + (a_{m+1}a_{m+1+k} + \cdots + a_{n-k}a_n) \\ &= S_k + (a_{m-k}a_{m+1} + \cdots + a_{m-1}a_{m+k}) - (a_{m-k}a_m + \cdots + a_ma_{m+k}). \end{aligned}$$

$$S_{k+1}' = S_{k+1} + (a_{m-k-1}a_{m+1} + \cdots + a_{m-1}a_{m+k+1}) - (a_{m-k-1}a_m + \cdots + a_ma_{m+k+1}).$$

When these are differenced we have, transferring the term $a_{m-k-1}a_m$,

$$\begin{aligned} S_k - S_{k+1} = (S_k' - S_{k+1}') &+ [(a_{m-k-1}a_{m+1} + \cdots + a_{m-1}a_{m+k+1}) \\ &- (a_{m-k-1}a_m + a_{m-k}a_{m+1} + \cdots + a_{m-1}a_{m+k})] \\ &+ [(a_{m-k}a_m + \cdots + a_ma_{m+k}) \\ &- (a_{m-k}a_{m+1} + \cdots + a_ma_{m+k+1})] \\ = (S_k' - S_{k+1}') &+ [a_{m-k}(a_m - a_{m+1}) + \cdots + a_m(a_{m+k} - a_{m+k+1})] \\ &- [a_{m-k-1}(a_m - a_{m+1}) + \cdots + a_{m-1}(a_{m+k} - a_{m+k+1})] \\ = (S_k' - S_{k+1}') &+ [(a_{m-k} - a_{m-k-1})(a_m - a_{m+1}) \\ &+ \cdots 2(a_m - a_{m-1})(a_{m+k} - a_{m+k+1})]. \end{aligned}$$

By the induction hypothesis, $S_k' - S_{k+1}'$ is nonnegative; by the unimodality assumption the term in square brackets is a sum of products of nonnegative terms. We conclude $S_k \geqq S_{k+1}$.

We can now establish the desired result.

THEOREM 2. *If X and Y are independent observations on the same unimodal random variable, then $X - Y$ is unimodal.*

We prove the theorem in three parts.

PART I. If X has as possible values only finitely many integers, the theorem is an immediate consequence of the preceding one. The a's are taken to be the probabilities of the successive possible values of X. Since $P(X - Y = k) = S_k$ for k a positive integer, and since $X - Y$ has a distribution symmetric about 0, the theorem follows.

PART II. Let the possible values of X now be numbers of the form $r\Delta$, where $\Delta > 0$ and r is any integer. For simplicity we may assume 0 to be a mode. For every positive integer s, define

$$X'_s = \begin{cases} X \text{ if } |X| \leqq s, \\ 0 \text{ if } |X| > s, \end{cases} \qquad Y'_s = \begin{cases} Y \text{ if } |Y| \leqq s \\ 0 \text{ if } |Y| > s. \end{cases}$$

That $X'_s - Y'_s$ has a unimodal distribution is an immediate consequence of Part I. But since $P(X'_s - Y'_s \neq X - Y) \to 0$ as $s \to \infty$, we see that $X - Y$ must also have a unimodal distribution.

PART III. Now suppose X has a density f, with mode at m. For each positive integer s, define

$$X''_s = [(X - m)\sqrt{s}]/\sqrt{s},$$

where $[u]$ denotes the greatest integer less than u. The cumulative distribution G''_s of $X''_s - Y''_s$ cannot ever differ from G by more than a quantity which tends to 0 as $s \to \infty$. However, G''_s is unimodal, by Part II. If G were not unimodal, we could find $\epsilon > 0$, $\Delta > 0$, and $u - \Delta > 0$ such that $G(u - \Delta) + G(u + \Delta) + \epsilon < 2G(u)$, which would yield a contradiction.

REFERENCE

[1] K. L. CHUNG, "Sur les lois de probabilité unimodales," *C. R. Acad. Sci. Paris*, Vol. 236 (1953), pp. 583–584.

NOTE ON A THEOREM OF LIONEL WEISS[1]

BY LUCIEN LECAM

University of California, Berkeley

1. Introduction. In a recent paper [1] it was pointed out by Lionel Weiss that the class of sequential probability ratio tests is complete in a very strong sense. The purpose of the present note is to show how this result can be derived from a

Received 2/8/54.

[1] This paper was prepared with the partial support of the Office of Naval Research.

slight extension of the usual theorems of decision theory, and to generalize this result to the case where the number of alternatives is any finite number. Similar results could also be obtained in more general cases.

2. Statement of the problem. Let $\{X_n\}$, for $n = 1, 2, \cdots$, be a sequence of mutually independent random vectors. Assume that the distribution of these vectors is given by a sequence of probability measures $\{\pi_{n,j}\}$, for $n = 1, 2, \cdots$, where j is an index taking one of the values $j = 1, 2, \cdots, k$. Suppose that the loss incurred by accepting j while i is true is a finite number W_{ij}, strictly positive if $i \neq j$ and equal to zero if $i = j$.

Let $\mathcal{W}$ be the class of matrices (W_{ij}) satisfying these conditions. If i is the true state of nature, we will assume that the cost of taking n observations is $C_i(n)$, nonnegative strictly increasing in n and such that $C_i(n)$ tends to infinity as n tends to infinity. Let $\mathcal{C}$ be the class of all k-tuples of cost functions $C = \{C_i(n)\}$ satisfying the preceding conditions. Let $\Theta = J \times \mathcal{W} \times \mathcal{C}$, where J denotes the set of integers $J = \{1, 2, \cdots, k\}$. For a particular decision function δ and a particular point $\theta = \{i, W, C\} \varepsilon \Theta$, let $R(\theta, \delta)$ denote the risk if δ is used for the state of nature i, the loss function W, and the cost function C.

If D is any subset of the set $\mathfrak{D}$ of all measurable decision procedures such that
(1) D contains all $\delta \varepsilon \mathfrak{D}$ which minimize linear combinations of the form

$$K(\mu, \delta) = \sum_{\xi=1}^{m} \mu_\xi K(\theta_\xi, \delta) \qquad \mu_\xi > 0; \quad \sum_{\xi=1}^{m} \mu_\xi = 1;$$

(2) D is compact in the sense defined in [2];
Then it follows from a modification ([2] Theorem (5)) of a theorem of Wald ([3] Theorem 3.18)) that D is essentially complete. This means that whatever $\delta_0 \varepsilon \mathfrak{D}$, there exists $\delta_1 \varepsilon D$ such that $R(\theta, \delta_1) \leqq R(\theta, \delta_0)$ for every $\theta \varepsilon \Theta$. Such an essentially complete class is described below.

3. Description of the complete class. Let Δ be the set of probability distributions on J. Let Z_n be the vector $Z_n = \{X_1 \cdots X_n\}$ and let $q(Z_{n,p}) = \{q_j(Z_{n,p})\}$ be the vector representing the a posteriori distribution of $i \varepsilon J$ given Z_n, when the a priori distribution of $i \varepsilon J$ is p. Let p be fixed. Consider the class D of all decision functions defined in the following way. For each $n \geqq 0$ choose k closed convex sets $\{S_{n,j}\}$ with $j = 1, 2, \cdots, k$, each contained in Δ, with disjoint interiors and such that $S_{n,j}$ contains q if $q = \{q_i\}$ with $q_j = 1$.

The decision function δ consists of the following rule: if $q(Z_{n,p}) \varepsilon S_{n,j}$, then stop and accept j; if $q(Z_{n,p})$ is not a member of $\bigcup_j S_{n,j}$, then take one more observation; if $q(Z_{n,p})$ is a limit point of one or many $S_{n,j}$, randomize appropriately.

It is clear that the preceding description uses characteristics not depending explicitly on p, so that p may be fixed and, for instance, taken equal to the uniform distribution on J. For this class D the following theorem holds.

THEOREM. *There exists on $\mathfrak{D}$ a topology $\mathfrak{I}$ for which* (1) $R(\theta, \delta)$ *is lower semicontinuous in δ for each $\theta \varepsilon \Theta$ and* (2) *$\mathfrak{D}$ is compact and D is a closed, compact, subset of $\mathfrak{D}$.*

If δ_0 is any measurable decision function $\delta \varepsilon \mathfrak{D}$, there exists a $\delta_1 \varepsilon D$ such that $R(\theta, \delta_0) \leqq R(\theta, \delta_1)$, whatever may be $\theta \varepsilon Z$.

PROOF. A topology $\mathfrak{I}$ having the desired properties has been defined more generally [2] by a process analogous to the one used by Wald [3] for the definition of regular convergence. A classical theorem (see [4], Vol. 1, p. 246; Vol. 2, p. 21; or [5]) states that the space of closed subsets of a compact metric space is compact for the usual definition of distance between sets. It then follows from the relationship between compactness in this sense and compactness in the sense of $\mathfrak{I}$ (or of regular convergence in [3]) that D is compact. This proves the first part of the theorem.

The second part is an immediate consequence of Theorem (5) in [2], provided we can show that Bayes' solutions belong to D. To show this, let $P_{ij}(\delta)$ be the probability of accepting j if i is true and δ is used, and let $Q_i(n, \delta)$ be the probability of taking at least n observations if i is true and δ is used. Let $\theta = \{i, W, C\}$. Then

$$R_i(W, C, \delta) = R(\theta, \delta) = \sum_j W_{ij} P_{ij}(\delta) + \sum_{n=0}^{\infty} C_i(n)[Q_i(n, \delta) - Q_i(n+1, \delta)].$$

Consequently,

$$\sum_{\xi=1}^{m} \mu_\xi R(\theta_\xi, \delta) = \sum_{i=1}^{k} p_i R_i(W, C, \delta)$$

for suitable values of W, C, and $\{p_i\}$. Therefore the Bayes' solutions for our problem have the same structure as the Bayes' solutions for the now classical problem in which W and C are fixed. A very slight modification of the argument given by Arrow, Blackwell, and Girshick [6] yields the desired result. This completes the proof of the theorem.

As a particular case, if $\delta_0 \varepsilon \mathfrak{D}$ is such that $\lim_{n\to\infty} Q_i(n, \delta_0) = 0$ for $i \varepsilon J$, the preceding theorem implies that there exists $\delta_1 \varepsilon D$ satisfying:

$$P_{ij}(\delta_0) \geqq P_{ij}(\delta_1), \qquad \text{for every } i, j \varepsilon J; \quad i \neq j;$$

$$Q_i(n, \delta_0) \geqq Q_i(n, \delta_1), \qquad \text{for every } i \varepsilon J; \quad n \geqq 0.$$

If, furthermore, the probabilities $\{\pi_{n,j}\}$ satisfy the condition imposed by Weiss [1], the boundaries of the S_{nj} have measure zero and randomization is unncessary.

4. Remarks.

(1) the technique of enlarging the space of strategies of nature, say Ω, to a product $\Omega \times S$ with S finite has been used systematically by Weiss [7], [8] and Lindley [9]. A more general type of extension is implicitly contained in the assumptions of [2]. The standard form of Bayes' solutions given by Theorem 4.7 of [3], or its generalization, remains usually valid under such modifications of Ω.

(2) The proofs of the optimum character of the sequential probability ratio test given in [6] or [10] also make use of classes of Bayes' solutions obtained by varying W and C. However, in these proofs C remains proportional to a given C_0. This is not sufficient for the present purpose.

REFERENCES

[1] LIONEL WEISS, "Testing one simple hypothesis against another," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 273–281.

[2] L. LECAM, "An extension of Wald's theory of statistical decision functions," unpublished.

[3] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, 1950.

[4] C. KURATOWSKI, "Topologie I," 2nd ed., Warsaw, 1948. "Topologie II," Warsaw, 1950.

[5] G. CHOQUET, "Convergences," Annales de l'université de Grenoble, N.S., Vol. 23 (1947–1948), pp. 57–112.

[6] K. J. ARROW, D. BLACKWELL AND M. A. GIRSHICK, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213–244.

[7] L. WEISS, "A higher order complete class theorem," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 677–680.

[8] L. WEISS, "Sequential procedures that control the individual probabilities of coming to the various decisions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 779–784.

[9] D. V. LINDLEY, "Statistical inference," *J. Roy. Stat. Soc. Series B*, Vol. 15, No. 1 (1953), pp. 30–76.

[10] A. WALD AND J. WOLFOWITZ, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–339.

THE DISTRIBUTION OF DISTANCE IN A HYPERSPHERE

BY R. D. LORD

The Royal Technical College, Glasgow

1. In a note with the above title, Hammersley [2] has used ad hoc methods to deal with the distribution of the distance AB, when A and B are points uniformly distributed in a sphere of radius a in s dimensions. I show here how this question may be treated by general methods which I have developed elsewhere [3] for random vectors with spherical distributions. A random vector $\mathbf{r}$ will be said to have a *spherical distribution* if its probability function is a function of $|\mathbf{r}|$ only.

I start with the observation that the problem is in fact one of the addition of independent random vectors with spherical distributions. We require the distribution of $\mathbf{r}_1 - \mathbf{r}_2$ where $\mathbf{r}_1$ and $\mathbf{r}_2$ are random vectors with the same uniform spherical distribution. But on account of the spherical symmetry, $-\mathbf{r}_2$ has the same distribution as $\mathbf{r}_2$, so that the problem is equivalent to finding the distribution of $\mathbf{r}_1 + \mathbf{r}_2$. It will be dealt with in this form in what follows.

2. The first method uses the polar form of the characteristic function. For any spherical distribution in s dimensions let

$$P(r)\,dr = Pr\{r < |\mathbf{r}| < r + dr\}.$$

Received 10/5/53, revised 4/12/54.

The characteristic function of the distribution of r is $E(e^{i\mathbf{r}\cdot\boldsymbol{\rho}})$. On changing to polar coordinates it is found (as in [1] or [3]) to be a function of $\rho = |\boldsymbol{\rho}|$ only, and is

$$\Phi(\rho) = \int_0^\infty P(r)\Lambda_{s/2-1}(r\rho)\,dr, \tag{1}$$

where

$$\begin{aligned}\Lambda_\alpha(x) &= \Gamma(\alpha+1)(\tfrac{1}{2}x)^{-\alpha}J_\alpha(x)\\ &= 1 - \frac{(\frac{1}{2}x)^2}{1\cdot(\alpha+1)} + \frac{(\frac{1}{2}x)^4}{1\cdot 2(\alpha+1)(\alpha+2)} - \cdots,\end{aligned} \tag{2}$$

with inversion formula

$$P(r) = 2^{-s/2+1}\{\Gamma(\tfrac{1}{2}s)\}^{-1}\int_0^\infty (r\rho)^{s/2}J_{s/2-1}(r\rho)\Phi(\rho)\,d\rho. \tag{3}$$

It should be emphasised that $\Phi(\rho)$ is the characteristic function of the s-dimensional distribution of $\mathbf{r}$ and not of the one-dimensional distribution of $r = |\mathbf{r}|$.

For a distribution uniform in a sphere of radius a

$$P(r) = \begin{cases} sr^{s-1}a^{-s}, & 0 \leq r \leq a;\\ 0, & r > a,\end{cases} \tag{4}$$

$$\Phi(\rho) = \Lambda_{s/2}(a\rho). \tag{5}$$

Multiplying characteristic functions and inverting, we obtain the probability function for $\mathbf{r}_1 + \mathbf{r}_2$ as

$$P_2(r) = s\Gamma(s/2+1)(2r/a^2)^{s/2}\int_0^\infty \rho^{-s/2}J^2_{s/2}(a\rho)J_{s/2-1}(r\rho)\,d\rho.$$

This integral is not completely evaluated by Watson [4], but we merely need to make simple substitutions (in line 6 of sec. 13.46 and in equation (2) of sec. 13.4). The result is

$$P_2(r) = \frac{s\Gamma(\frac{1}{2}s+1)}{\Gamma(\frac{1}{2}s+\frac{1}{2})\Gamma(\frac{1}{2})}r^{s-1}a^{-s}\int_A^\pi \cos^s \tfrac{1}{2}\phi\,d\phi,$$

where $0 \leq A \leq \pi$ and $\sin\frac{1}{2}A = r/2a$. Putting $t = \cos^2\frac{1}{2}\phi$, we obtain Hammersley's form

$$P_2(r) = sr^{s-1}a^{-s}I_\mu(\tfrac{1}{2}s+\tfrac{1}{2},\tfrac{1}{2}), \tag{6}$$

where $\mu = 1 - r^2/4a^2$ and $I_x(p, q)$ is the incomplete Beta function defined by

$$B(p, q)I_x(p, q) = \int_0^x t^{p-1}(1-t)^{q-1}\,dt.$$

3. In the second method the distributions are treated as projections of spherical distributions in a space of a higher number of dimensions. It is clear that

from a spherical distribution of random vectors with O for center (i.e. O is the point $\mathbf{r} = 0$) we obtain another spherical distribution with center O if we project the vectors orthogonally onto a space of lower dimensions through O. A simple calculation [3] shows that if any spherical distribution in space of $(s + 2m)$ dimensions is projected onto a space of s dimensions, the corresponding probability functions $P^{(s+2m)}(r)$ and $P^{(s)}(r)$, satisfy

$$P^{(s)}(r) = \frac{2\Gamma(\frac{1}{2}s + m)}{\Gamma(\frac{1}{2}s)\Gamma(m)} r^{s-1} \int_r^\infty P^{(s+2m)}(t)(t^2 - r^2)^{m-1} t^{-s-2m+2}\, dt. \tag{7}$$

If the distribution in the higher space is uniform over the surface of a sphere of radius a, then

$$P^{(s)}(r) = \begin{cases} \dfrac{2\Gamma(\frac{1}{2}s+m)}{\Gamma(\frac{1}{2}s)\Gamma(m)} a^{-s-2m+2}(a^2 - r^2)^{m-1} r^{s-1}, & r \leqq a, \\ 0, & r > a. \end{cases} \tag{8}$$

When $m = 1$, this reduces to (4).

This shows that a uniform distribution through the volume of an s-dimensional sphere can be obtained by projection from a uniform distribution over the surface of an $(s + 2)$-dimensional sphere, each sphere having radius a. In the case $s = 1$, we see that a distribution uniform over a diameter can be obtained by projection from a distribution uniform over the surface of a sphere. This is essentially Archimedes' theorem on the surface area of a sphere.

Now for the sum of two vectors, each with a distribution uniform over the surface of an $(s + 2)$-dimensional sphere, we can appeal to a special case of Kluyver's original solution of the problem of random flights, or rather to the generalisation to any number of dimensions given by Watson [4]. From his results ([4], secs. 13.48 and 13.46 (3)), it follows that the probability function is

$$P_2^{(s+2)}(r) = \begin{cases} 2^{1-s} \dfrac{\Gamma(\frac{1}{2}s + 1)}{\Gamma(\frac{1}{2}s + \frac{1}{2})\Gamma(\frac{1}{2})} a^{-2s} r^s (4a^2 - r^2)^{(s-1)/2}, & r \leqq 2a; \\ 0, & r > 2a. \end{cases} \tag{9}$$

Substituting in (7) with $m = 1$, we obtain $P_2(r)$ as a multiple of $\int_r^{2a} (4a^2 - t^2)^{(s-1)/2}\, dt$, and then (6) follows.

If the distribution of each $\mathbf{r}_1$ and $\mathbf{r}_2$ is according to (8), with m not necessarily equal to 1, then it is the projection from space of dimensions $(s + 2m)$ of a distribution uniform over the surface of a sphere. The argument just used is applicable and shows that

$$P_2(r) = k \int_r^{2a} (4a^2 - t^2)^{m+(s-3)/2} (t^2 - r^2)^{m-1} t^{-2m+2}\, dt,$$

where k is a constant. When $m = 1$ this reduces to (4), but is otherwise more complicated. The result is still true when $2m$ is not an integer.

4. Hammersley proves that for large values of s the distance between two points in the sphere is nearly always equal to $a\sqrt{2}$, the diagonal of the rectangle determined by orthogonal radii. He does this by showing that as s tends to infinity, $|\mathbf{r}_1 + \mathbf{r}_2|$ is asymptotically distributed in a normal distribution with mean $a\sqrt{2}$ and variance $a^2/2s$.

From the characteristic function it is seen that Hammersley's result is a corollary of a more general one, namely that the s-dimensional distribution given by (8) is asymptotically normal with second moment $a^2s(s + 2m)^{-1}$. Here a normal distribution has the probability function

$$P(r) = C_s r^{s-1} \exp\left(-\tfrac{1}{2}sr^2/\mu_2\right),$$

where μ_2 is the second moment and C_s a constant, and has the characteristic function

$$\Phi(\rho) = \exp\left(-\tfrac{1}{2}\mu_2\rho^2/s\right).$$

The distribution (8) has characteristic function $\Lambda_{s/2+m-1}(a\rho)$. This can be verified by direct calculation, or derived from the facts that a spherical distribution and its projections (in the sense of sec. 3) all have the same characteristic function (proved in [3]), and that a distribution uniform over the surface of a sphere of radius a in $s + 2m$ dimensions obviously has the characteristic function $\Lambda_{s/2+m-1}(a\rho)$. Now

$$\begin{aligned}\Lambda_{s/2+m-1}(a\rho) &= 1 - \frac{a^2\rho^2}{2(s + 2m)} + \frac{a^4\rho^4}{8(s + 2m)(s + 2m + 2)} - \cdots \\ &\sim \exp\left\{-\frac{a^2\rho^2}{2(s + 2m)}\right\}\end{aligned}$$

as s tends to infinity, uniformly in any ρ-interval. Thus the distribution (8) is asymptotically normal with $\mu_2 = a^2s(s + 2m)^{-1}$.

Taking $m = 1$, we obtain the distribution (4) which is therefore asymptotically normal with $\mu_2 = a^2s(s + 2)^{-1}$. The distribution of $\mathbf{r}_1 + \mathbf{r}_2$ is thus asymptotically normal with

$$\mu_2 = 2a^2s(s + 2)^{-1}. \tag{10}$$

Taking $m = 0$, we see that the distribution uniform over the surface of a sphere of radius b is asymptotic to a normal distribution with $\mu_2 = b^2$. Comparing with (10), we see that the distribution of $\mathbf{r}_1 + \mathbf{r}_2$ is asymptotic to a distribution uniform over the surface of a sphere of radius $a(2s)^{1/2}(s + 2)^{-1/2} \simeq a\sqrt{2}(1 + s^{-1})$. This is equivalent to Hammersley's result.

We could avoid the use of characteristic functions in an increasing number of dimensions by projecting onto a diametral subspace of a fixed number of dimensions. Since projection does not alter the characteristic function, the resulting calculation will be the same.

REFERENCES

[1] S. Bochner and K. Chandrasekharan, "Fourier transforms," *Ann. Math. Studies*, No. 19 (1949), Princeton.

[2] J. M. Hammersley, "The distribution of distance in a hypersphere," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 447–452.

[3] R. D. Lord, "The use of the Hankel transform in statistics. I, General theory and examples," to be published in *Biometrika*.

[4] G. N. Watson, *A treatise on Bessel functions*, Cambridge University Press, 1st edn., 1922, 2nd edn., 1944.

EXTREME VALUES IN SAMPLES FROM m-DEPENDENT STATIONARY STOCHASTIC PROCESSES

By G. S. Watson

University of Melbourne, Australia

Summary. The limiting distributions for the order statistics of n successive observations in a sequence of independent and identically distributed random variables are shown to hold also when the sequence is generated by a stationary stochastic process of a certain moving average type.

A sequence of random variables $\{x_i\}$ has been called m-dependent [3] if $|i - j| > m$ implies that x_i and x_j are independent. If the variables in a strictly stationary sequence are m-dependent and have a finite upper bound to their range of variation, the largest in a sample of n successive members tends with probability one to this upper bound. This is a simple extension of Dodd's results [1] for the case of independence.

The following theorem shows that when this upper bound is infinite, the asymptotic distribution of the largest in such a sample is the same as in the case of independence.

Theorem. *Let $\{x_i\}$ be a sequence of random variables, unbounded above and generated by an m-dependent strictly stationary stochastic process with the property that*

$$\lim_{c\to\infty} \frac{1}{P(x_i > c)} \max_{|i-j|\leq m} P[(x_i > c), (x_j > c)] = 0. \tag{1}$$

Then, if $\xi = n\, P[x_i > c_n(\xi)]$, for ξ fixed,

$$\lim_{n\to\infty} P[x_i \leq c_n(\xi); i = 1, \cdots, n] = e^{-\xi}$$

Proof. Using the formula for the probabilities of the joint occurrence of a set of events in terms of probabilities of occurrence of their contraries (Feller [2],

Received 8/19/52, revised 7/2/54.

p. 61), we have, for any even integer $l \leqq n$ and for $i = 1, \cdots, n$, that $P[x_i \leqq c_n(\xi)]$ is bounded below and above, respectively, by

$$1 - \sum P(x_i > c) + \cdots + (-1)^{l-1} \sum P[(x_{i_1} > c), \cdots, (x_{i_{l-1}} > c)],$$

$$1 - \sum P(x_i > c) + \cdots + (-1)^{l} \sum P[(x_{i_1} > c), \cdots, (x_{i_l} > c)],$$

where, for brevity, $c = c_n(\xi)$. Clearly, $\sum P(x_i > c) = nP(x_i > c) = \xi$.

Now

$$\sum P[(x_i > c), (x_j > c)]$$
$$= \sum_{i=1}^{m} (n - i)P[(x_1 > c), (x_{i+1} > c)] + P(x_i > c)^2 \tfrac{1}{2}(n - m - 1)(n - m).$$

But

$$\sum_{i=1}^{m} (n - i)P[(x_1 > c), (x_{i+1} > c)]$$
$$\leqq mn\left[1 - \frac{(m + 1)}{2n}\right] \max_{|i-j| \leqq m} P[(x_i > c), (x_j > c)]$$
$$= m\xi\left[1 - \frac{m + 1}{2n}\right] \max_{|i-j| \leqq m} \frac{P[(x_i > c), (x_j > c)]}{P(x_i > c)}.$$

Since, as $n \to \infty$, with ξ fixed, $c = c_n(\xi) \to \infty$, condition (1) shows that the last expression tends to zero. Hence

$$\lim_{n \to \infty} \sum P[(x_i > c), (x_j > c) = \tfrac{1}{2}\xi^2.$$

The general sum $\sum P[(x_{i_1} > c), \cdots, (x_{i_q} > c)]$ contains $\binom{n}{q}$ terms. Of these, there are order n terms in which none of the x_i appearing ever differs in its subscript by more than m from its nearest neighbours, order n^2 terms in which only one x_i differs in its subscript by more than m from its nearest neighbours, and so on, provided that n is large enough for all the cases to occur. There are $\sim n^q/q!$ terms in which each x_i is separated in its subscript by more than m from its neighbours. These terms may be said to belong to the first, second, $\cdots$, qth class. The sum of terms of the first class will be less than a constant times $n \max_{|i-j| \leqq m} P[(x_i > c), (x_j > c)]$, the sum of terms of the second class will be less than a constant times $n^2 P(x_i > c) \max_{|i-j| \leqq m} P[(x_i > c), (x_j > c)]$ and so on until we reach the qth class, where the sum is $[n^q/q! + O(n^{q-1})]\, P(x_i > c)^q$. Thus, by (1), the only terms contributing to the sum as $n \to \infty$ are those of the qth class; these yield $\xi^q/q!$ asymptotically.

Thus for any even integer l we have shown that

$$\sum_{q=0}^{l-1} \frac{(-\xi)^q}{q!} \leqq \lim_{n \to \infty} P(x_i \leqq c_n(\xi) \leqq \sum_{q=0}^{l} \frac{(-\xi)^q}{q!}, \qquad i = 1, \cdots, n$$

which proves the theorem.

To show that this theorem covers a class of stochastic processes of practical interest, it is shown next that the condition (1) of the theorem is true in strictly stationary processes which are normal. For this, it suffices to show that

$$\frac{P[(x > c), (y > c)]}{P(x > c)} \to 0, \qquad (c \to \infty), \tag{2}$$

where x and y have a bivariate normal distribution with means zero, variances unity and covariance ρ, with $|\rho| < 1$. Now

$$P[(x > c), (y > c)]$$
$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_c^\infty \int_c^\infty \exp\left[\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] dx\, dy.$$

The substitution $x = r/c + c$ and $y = t/c + c$ leads to

$$P[(x > c), (y > c)]$$
$$= \frac{\exp[(-c^2/(1+\rho)]}{2\pi c^2\sqrt{1-\rho^2}} \int_0^\infty \int_0^\infty \exp\left[-\frac{r^2 - 2\rho rt + t^2}{2c^2(1-\rho^2)}\right] \exp\left[-\frac{r+t}{1+\rho}\right] dr\, dt$$
$$\sim \frac{1}{2\pi} \exp\left(\frac{-c^2}{1+\rho}\right)\left[\frac{(1+\rho)^{3/2}}{\sqrt{1-\rho}}\frac{1}{c^2} - 0\left(\frac{1}{c^4}\right)\right], \qquad c \text{ large.}$$

Since $P(x > c) \sim (1/\sqrt{2\pi}) \exp(-\frac{1}{2}c^2)$, statement (2) follows.

Acknowledgement. The author wishes to record his gratitude to a referee for suggesting a change in the assumptions made in an earlier form in this paper.

REFERENCES

[1] E. L. Dodd, "The greatest and the least variate under general laws of error," *Trans. Amer. Math. Soc.*, Vol. 25 (1923), p. 525–539.

[2] W. Feller, *An introduction to probability theory and its applications*, John Wiley and Sons, Inc., New York., 1950.

[3] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math. J.*, Vol. 15 (1948), pp. 773–780.

EXPRESSION OF THE k-STATISTICS k_9 AND k_{10} IN TERMS OF POWER SUMS AND SAMPLE MOMENTS

By M. Zia ud-Din

Panjab University, Lahore, Pakistan

The k statistics are of interest to workers in the theory of sampling distributions and moment statistics. They are related also to certain aspects of the theory of numbers and combinatory analysis, as indicated by Dressel [1].

The k statistics were introduced by Fisher in 1928 [2] to estimate the cumulants

Received 8/25/53, revised 3/11/54.

or Thiele semi-invariants of a population. Dressel [1] has given a table of the k_r $(r = 1, 2, \cdots, 8)$ in terms of the sums s_r of the rth powers of the observations in a sample of size n. This note adds k_9 and k_{10} to those available in print.

One may readily obtain k_9 and k_{10} in terms of the sample moments m_r about the sample mean by replacing s_1 by zero and s_r $(r > 1)$ by nm_r in the following expressions.

The expressions for k_9 and k_{10} were obtained by following Kendall [3] and using tables of the symmetric functions [4]. The work has been carefully checked. A fundamental check given by Dressel has been successfully applied to both expressions. This check has revealed a correction for L_6 as given by Dressel: the coefficient of (5) (1) in L_6 should be

$$-15(n^4 + 2n^3 - 7n^2 + 4n).$$

It is found that $n^{(9)}k_9$

$$
\begin{aligned}
&= (n^8 + 219n^7 + 3721n^6 + 6189n^5 - 7250n^4 + 2160n^3)s_9 \\
&\quad - 9(n^7 + 219n^6 + 3721n^5 + 6189n^4 - 7250n^3 + 2160n^2)\, s_8s_1 \\
&\quad - 36(n^7 + 93n^6 + 277n^5 - 1917n^4 + 2746n^3 - 1200n^2)s_7s_2 \\
&\quad + 72(n^6 + 156n^5 + 1999n^4 + 2136n^3 - 2252n^2 + 480n)s_7s_1^2 \\
&\quad - 84(n^7 + 33n^6 - 83n^5 + 543n^4 - 1214n^3 + 720n^2)s_6s_3 \\
&\quad + 504(n^6 + 63n^5 + 97n^4 - 687n^3 + 766n^2 - 240n)s_6s_2s_1 \\
&\quad - 504(n^5 + 94n^4 + 731n^3 + 254n^2 - 240n)s_6s_1^3 \\
&\quad - 126(n^7 + 9n^6 + 61n^5 - 201n^4 + 370n^3 - 240n^2)s_5s_4 \\
&\quad + 1008(n^6 + 21n^5 - 11n^4 + 171n^3 - 422n^2 + 240n)s_5s_3s_1 \\
&\quad + 756(n^6 + 18n^5 - 113n^4 + 198n^3 - 104n^2)s_5s_2^2 \\
&\quad - 4536(n^5 + 34n^4 - 9n^3 - 106n^2 + 80n)s_5s_2s_1^2 \\
&\quad + 3024(n^4 + 49n^3 + 176n^2 - 16n)s_5s_1^4 \\
&\quad + 630(n^6 + 9n^5 + 61n^4 - 201n^3 + 370n^2 - 240n)s_4^2s_1 \\
&\quad + 2520(n^6 - 5n^4 + 4n^2)s_4s_3s_2 \\
&\quad - 7560(n^5 + 10n^4 + 15n^3 - 10n^2 - 16n)s_4s_3s_1^2 \\
&\quad - 11340(n^5 + 6n^4 - 41n^3 + 66n^2 - 32n)s_4s_2^2s_1 \\
&\quad + 30240(n^4 + 14n^3 - 19n^2 + 4n)s_4s_2s_1^3 \\
&\quad - 15120(n^3 + 21n^2 + 20n)s_4s_1^5 \\
&\quad + 560(n^6 - 6n^5 + 31n^4 - 66n^3 + 40n^2)s_3^3 \\
&\quad - 15120(n^5 - 2n^4 + 7n^3 - 22n^2 + 16n)s_3^2s_2s_1 \\
&\quad + 20160(n^4 + 4n^3 + 11n^2 - 16n)s_3^2s_1^3 \\
&\quad - 7560(n^5 - 6n^4 + 11n^3 - 6n^2)s_3s_2^3 \\
&\quad + 90720(n^4 - n^3 - 4n^2 + 4n)s_3s_2^2s_1^2 \\
&\quad - 151200(n^3 + 3n^2 - 4n)s_3s_2s_1^4 \\
&\quad + 60480(n^2 + 6n)s_3s_1^6 + 22680(n^4 - 6n^3 + 11n^2 - 6n)s_2^4s_1 \\
&\quad - 151200(n^3 - 3n^2 + 2n)s_2^3s_1^3 + 272160(n^2 - n)s_2^2s_1^5 \\
&\quad - 181440n\, s_2s_1^7 + 40320\, s_1^9
\end{aligned}
$$

Similarly it is found that $n^{(10)}k_{10}$

$$
\begin{aligned}
= {} & (n^9 + 466n^8 + 15706n^7 + 72976n^6 - 41171n^5 - 41186n^4 + 45624n^3 - 12096n^2)s_{10} \\
& - 10(n^8 + 466n^7 + 15706n^6 + 72976n^5 - 41171n^4 - 41186n^3 + 45624n^2 - 12096n)s_9s_1 \\
& - 45(n^8 + 212n^7 + 2428n^6 - 9166n^5 + 859n^4 + 27098n^3 - 33528n^2 + 12096n)s_8s_2 \\
& + 90(n^7 + 339n^6 + 9067n^5 + 31905n^4 - 20156n^3 - 7044n^2 + 6048n)s_8s_1^2 \\
& - 120(n^8 + 88n^7 + 40n^6 + 526n^5 + 2719n^4 - 18758n^3 + 27480n^2 - 12096n)s_7s_3 \\
& + 720(n^7 + 150n^6 + 1234n^5 - 4320n^4 + 1789n^3 + 4170n^2 - 3024n)s_7s_2s_1 \\
& - 720(n^6 + 213n^5 + 3845n^4 + 7755n^3 - 5526n^2 + 432n)s_7s_1^3 \\
& - 210(n^8 + 32n^7 + 88n^6 + 734n^5 - 5441n^4 + 17378n^3 - 24888n^2 + 12096n)s_6s_4 \\
& + 1680(n^7 + 60n^6 + 64n^5 + 630n^4 - 1361n^3 - 690n^2 + 1296n)s_6s_3s_1 \\
& + 1260(n^7 + 57n^6 - 203n^5 - 465n^4 + 2794n^3 - 3912n^2 + 1728n)s_6s_2^2 \\
& - 7560(n^6 + 89n^5 + 365n^4 - 1385n^3 + 1074n^2 - 144n)s_6s_2s_1^2 \\
& + 5040(n^5 + 120n^4 + 1235n^3 + 900n^2 - 576n)s_6s_1^4 \\
& - 126(n^8 + 16n^7 + 256n^6 - 1274n^5 + 5959n^4 - 16886n^3 + 24024n^2 - 12096n)s_5^2 \\
& + 2520(n^7 + 24n^6 + 172n^5 - 270n^4 + 259n^3 + 246n^2 - 432n)s_5s_4s_1 \\
& + 5040(n^7 + 15n^6 - 101n^5 + 405n^4 - 1196n^3 + 1740n^2 - 864n)s_5s_3s_2 \\
& - 15120(n^6 + 33n^5 + 45n^4 + 255n^3 - 766n^2 + 432n)s_5s_3s_1^2 \\
& - 22680(n^6 + 29n^5 - 135n^4 + 115n^3 + 134n^2 - 144n)s_5s_2^2s_1 \\
& + 60480(n^5 + 45n^4 + 35n^3 - 225n^2 + 144n)s_5s_2s_1^3 \\
& - 30240(n^4 + 60n^3 + 275n^2)s_5s_1^5 \\
& + 3150(n^7 + 3n^6 + 31n^5 - 375n^4 + 1264n^3 - 1788n^2 + 864n)s_4^2s_2 \\
& - 9450(n^6 + 17n^5 + 125n^4 - 305n^3 + 594n^2 - 432n)s_4^2s_1 \\
& + 4200(n^7 - 3n^6 + 25n^5 - 45n^4 - 26n^3 + 48n^2)s_4s_3^2 \\
& - 75600(n^6 + 5n^5 - 15n^4 - 5n^3 + 14n^2)s_4s_3s_2s_1 \\
& + 100800(n^5 + 15n^4 + 35n^3 - 15n^2 - 36n)s_4s_3s_1^3 \\
& - 18900(n^6 + n^5 - 55n^4 + 215n^3 - 306n^2 + 144n)s_4s_2^3 \\
& + 226800(n^5 + 10n^4 - 55n^3 + 80n^2 - 36n)s_4s_2^2s_1^2 \\
& - 378000(n^4 + 18n^3 - 19n^2)s_4s_2s_1^4 \\
& + 151200(n^3 + 25n^2 + 30n)s_4s_1^6 \\
& - 16800(n^6 - 3n^5 + 25n^4 - 45n^3 - 26n^2 + 48n)s_3^3s_1 \\
& - 37800(n^6 - 7n^5 + 25n^4 + 65n^3 + 94n^2 - 48n)s_3^2s_2^2 \\
& + 302400(n^5 + 5n^3 - 30n^2 + 24n)s_3^2s_2s_1^2 \\
& - 252000(n^4 + 6n^3 + 17n^2 - 24n)s_3^2s_1^4 \\
& + 302400(n^5 - 5n^4 + 5n^3 + 5n^2 - 6n)s_3s_2^3s_1 \\
& - 1512000(n^4 - 7n^3 + 6n)s_3s_2^2s_1^3 \\
& + 1814400(n^3 + 4n^2 - 5n)s_3s_2s_1^5 \\
& - 604800(n^2 + 7n)s_3s_1^7 \\
& + 22680(n^5 - 10n^4 + 35n^3 - 50n^2 + 24n)s_2^5
\end{aligned}
$$

$$
\begin{aligned}
&- 567000(n^4 - 6n^3 + 11n^2 - 6n)s_2^4 s_1^2 \\
&+ 2268000(n^3 - 3n^2 + 2n)s_2^3 s_1^4 \\
&- 3175200(n^2 - n)s_2^2 s_1^6 \\
&+ 1814400n\, s_2 s_1^8 \\
&- 362880\, s_1^{10}
\end{aligned}
$$

REFERENCES

[1] P. L. DRESSEL, "Statistical seminvariants and their estimates with particular emphasis on their relation to algebraic invariants," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 33–57.

[2] R. A. FISHER, "Moments and product moments of sampling distributions," *Proc. London Math. Soc.*, Series 2, Vol. 30 (1928), pp. 199–238.

[3] M. G. KENDALL, *Advanced Theory of Statistics*, Vol. 1, Charles Griffin and Co., London, 1952, pp. 254–261.

[4] F. N. DAVID AND M. G. KENDALL, "Tables of symmetric functions, I." *Biometrika*, Vol. 36 (1949), pp. 431–449.

THE PROBABILITY INTEGRAL OF RANGE FOR SAMPLES FROM A SYMMETRICAL UNIMODAL POPULATION

BY J. H. CADWELL

Ordnance Board,[1] Great Britain

1. Summary. An asymptotic expression is given for the probability integral of range for samples from a symmetrical unimodal population. Its accuracy is investigated for the case of a normal parent population and for sample sizes from 20 to 100. Over this range errors are small, and by using a correction based on values given below the probability integral can be found with a maximum error of 0.0001. Percentage points of range in the normal case are tabled for $n = 20, 40, 60, 80$ and 100.

2. The asymptotic expansion. The parent probability density function $\phi(x)$ is symmetrical about $x = 0$ and its integral from 0 to x is denoted by $\Phi(x)$. The p.d.f. of w, the range for a sample of size n, is

$$p(w) = n(n-1) \int_{-\infty}^{\infty} \{\Phi(x) - \Phi(x-w)\}^{n-2} \phi(x)\phi(x-w)\, dx. \tag{1}$$

Integrating with respect to w from $-\infty$ to w gives

$$F(w) = n \int_{-\infty}^{\infty} \{\Phi(x) - \Phi(x-w)\}^{n-1} \phi(x)\, dx. \tag{2}$$

Received 11/10/53, revised 4/16/54.

[1] This work arose during analyses made by the Board's Statistics Group.

Hartley [2] proves that this can be transformed to

$$(3)\quad F(2u) = \{2\Phi(u)\}^n + 2n \int_0^\infty \{\Phi(x+u) - \Phi(x-u)\}^{n-1}\phi(x+u)\,dx.$$

Since $\phi(x)$ is unimodal the integrand in (3) is greatest when $x = 0$ and decreases rapidly to zero on either side of this point. This suggests the application of the methods used [1] to furnish asymptotic series for similar integrals. We have

$$\Phi(x+u) - \Phi(x-u) = 2\Phi(u)\{1 + A(u)x^4 + \cdots\} \exp\left\{\frac{x^2\phi'(u)}{2\Phi(u)}\right\},$$

$$\phi(x+u) = \phi(u)\{1 + B(u)x^3 + \cdots\} \exp\left\{\frac{x\phi'(u)}{\phi(u)} - \frac{x^2}{2}\left(\frac{\phi'(u)}{\phi(u)}\right)^2 + \frac{x^2\phi''(u)}{2\phi(u)}\right\}.$$

Ignoring all but the first term in each series, we substitute in (3) to obtain

$$(4)\quad \begin{aligned} F(2u) \simeq \{2\Phi(u)\}^n + 2\sqrt{2\pi}\, nk\phi(u)\{2\Phi(u)\}^{n-1}\{\tfrac{1}{2} - \Phi(-k\phi'(u)/\phi(u))\} \\ \exp\{\tfrac{1}{2}k^2(\phi'(u)/\phi(u))^2\}, \end{aligned}$$

where

$$k^{-2} = (\phi'(u)/\phi(u))^2 - (\phi''(u)/\phi(u)) - (n-1)\phi'(u)/\Phi(u).$$

When $\phi(x) = \exp - \frac{1}{2}x^2/\sqrt{2\pi}$, we find that (4) reduces to

$$(5)\quad F(2u) \simeq \{2\Phi(u)\}^n + 2nk\{2\Phi(u)\}^{n-1}\{\exp - \tfrac{1}{2}u^2(1-k^2)\}\{\tfrac{1}{2} - \Phi(uk)\},$$

where

$$k^{-2} = 1 + (n-1)u\phi(u)/\Phi(u).$$

In this case it is easy to include a further term, which results in the last bracket of (5) being replaced by

$$(6)\quad \{\tfrac{1}{2} - \Phi(uk) - (n-1)k^4P(u)Q(uk)\},$$

where

$$P(x) = \frac{x^2}{8}\left(\frac{\phi(x)}{\Phi(x)}\right)^2 + \frac{x^3 - 3x}{24}\frac{\phi(x)}{\Phi(x)},$$

$$Q(x) = (x^4 + 6x^2 + 3)\{\tfrac{1}{2} - \Phi(x)\} - (x^3 + 5x)\phi(x).$$

3. Accuracy in the normal case. While w is not defined when $n = 1$, expression (2) gives $F(w)$ the formal value unity for all w. This is also the value given by (5) or (6). Thus our expression, besides being asymptotically correct, also gives the exact value when $n = 1$. Hence [1], errors will at first rise with increase of n and then fall asymptotically to zero.

The following values of maximum error for (5) and (6) are the differences between exact values obtained by evaluating the p.d.f. using (1) and values of

$F(w)$ then found by quadrature.

Sample size, n:	20	60	100
Maximum error of (5):	+0.0031	+0.0040	+0.0043
Maximum error of (6):	−0.00052	−0.00070	−0.00075

By using (6), results of reasonable accuracy are obtained. Table I gives corrections in units in the fourth decimal place to be added to the approximate value given by (6), for five sample sizes. The corrections are given as functions of the approximate value itself, rather than of w, to make interpolation for n much simpler. By plotting the correction against the approximate probability on

TABLE I

Corrections ($\times 10^4$) *to be applied to approximate value obtained from equation* (6), *for samples of size n*

n	Value obtained from equation (6)									
	.05	.10	.25	.50	.75	.90	.95	.99	.995	.999
20	0	0.1	0.4	1.7	4.2	5.1	4.1	1.5	0.8	0.2
40	0.1	0.3	0.8	2.6	5.3	6.2	4.9	1.6	0.8	0.2
60	0.2	0.4	1.0	3.1	6.1	6.9	5.5	1.6	0.8	0.2
80	0.2	0.5	1.2	3.5	6.6	7.2	5.9	1.6	0.8	0.2
100	0.2	0.5	1.2	3.6	6.9	7.4	6.1	1.6	0.8	0.2

TABLE II

Percentage points of range (w) *for samples of various sizes* (n) *from normal populations of unit standard deviation*

n	Percentage points													Mean $\bar{w}$
	.001	.005	.010	.050	.100	.250	.500	.750	.900	.950	.990	.995	.999	
20	1.88	2.13	2.25	2.63	2.84	3.22	3.69	4.20	4.69	5.01	5.65	5.89	6.40	3.73
40	2.62	2.85	2.97	3.31	3.50	3.85	4.27	4.74	5.20	5.50	6.09	6.32	6.81	4.32
60	3.03	3.24	3.35	3.68	3.86	4.19	4.59	5.04	5.48	5.76	6.34	6.55	7.04	4.64
80	3.30	3.50	3.61	3.92	4.10	4.42	4.81	5.24	5.67	5.95	6.51	6.73	7.20	4.85
100	3.50	3.71	3.80	4.11	4.28	4.59	4.97	5.39	5.81	6.09	6.64	6.85	7.31	5.02

arithmetical probability paper, we can interpolate graphically for n and the approximate value. This will enable the probability integral to be found with an error that should not exceed 0.0001, and will usually be less than 0.00005.

Table II, giving percentage points of range found by quadrature, will assist in making preliminary estimates. Plotting $(w - \bar{w})$ against $n^{-1/2}$ for a given percentage level, permits interpolation for other values of n to be made with accuracy. Values of $\bar{w}$ are tabled in [3].

Acknowledgement. The author is indebted to the Chief Scientist, British Ministry of Supply, for permission to publish this note.

REFERENCES

[1] J. H. Cadwell, "The distribution of quasi-ranges in samples from a normal population," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 603–613.

[2] H. O. Hartley, "The range in normal samples," *Biometrika*, Vol. 32 (1942), pp. 334–348.

[3] K. Pearson, *Tables for Statisticians and Biometricians*, Part II, Biometrika Office, University College, London, 1931.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Montreal meeting of the Institute, September 10–13, 1954)

1. On Quadratic Estimates of Variance Components in Balanced Models, A. W. Wortham, Chance Vought Aircraft and Oklahoma A and M College.

A balanced model is defined as a model whose analysis of variance mean squares are symmetric in the squares of the observations. Included in this class of models are: (1) Completely Randomized, (2) Randomized Blocks, (3) Latin Squares, (4) Graeco-Latin Squares, (5) Split Plots, (6) Factorial Arrangements, etc.

The "analysis of variance estimates" of the variance components are the estimates obtained by solving the system of equations which result when the observed and expected mean squares in the analysis of variance table are equated. For any infinite population let the general balanced model be $y_{i_1 i_2 \cdots i_n} = \mu + \sum_{k=1}^{n} A_{ki_k} + e_{i_1 i_2 \cdots i_n}$, where μ is a constant, A_{ki_k} and $e_{i_1 i_2 \cdots i_n}$ are independent random variables with zero means, finite fourth moments, and variances σ_k^2 and σ_0^2 respectively. Let $\hat{\sigma}_k^2$ and $\hat{\sigma}_0^2$ be "the analysis of variance estimates" of the variance components σ_k^2 and σ_0^2. It is shown that the quadratic estimate of $\sum_{k=0}^{p} g_k \sigma_k^2$ (g_k known) which is unbiased, independent of μ, and has minimum variance is given by $\sum_{k=0}^{p} g_k \hat{\sigma}_k^2$. That is, the best quadratic unbiased estimate of the linear combination of the variance components is given by the same linear combination of "the analysis of variance estimates" of the variance components.

2. The Coefficients in the Best Linear Estimate of the Mean in Symmetric Populations, A. E. Sarhan, University of North Carolina.

In a previous paper ("Estimation of the Mean and Standard Deviation by Order Statistics" by A. E. Sarhan, *Ann. Math. Stat.* Vol. 25 (1954), pp. 317–328) the best linear estimate of the mean of a rectangular, triangular and double exponential population were worked out. By considering some other symmetric distributions with different shapes, it is found that the coefficients in the estimates form a sequence. From the sequence, it is observed that the coefficients in the estimates are influenced by the shape of the distribution. The variances of the estimates are also so affected.

3. Distribution of Linear Contrasts of Order Statistics, Jacques St. Pierre, University of North Carolina.

Consider $n + 1$ independent normal populations with unknown means, $m_0, m_1, \cdots, m_n$, respectively, and with a common known variance $\sigma^2 = 1$ (say). Suppose a sample of size N is available from each population; and let $x_{(0)} > x_{(1)} > \cdots > x_{(n)}$ be the ordered sample means. Consider the linear contrasts $z = x_{(0)} - c_1 x_{(1)} - \cdots - c_n x_{(n)}$, where $\sum_{i=1}^{n} c_i = 1$, $c_i \geq 0$, $(i = 1, 2, \cdots, n)$. The probability density function of the contrasts z is derived under the null hypothesis $H_0: m_0 = m_1 = \cdots = m_n$. The density of the contrasts z is also

obtained in the case of three populations, under the hypothesis $-\infty < m_2 \leq m_1 \leq m_0 < +\infty$. Particular hypotheses are considered and tables are given. Finally, the particular contrast $y = x_{(0)} - x_{(1)}$ is considered in the general case.

4. Note on Fourier Periodogram Analyses of Time Series, B. F. Kimball, New York State Public Service Commission.

R. A. Fisher's treatment of the probability distribution of the squares of the amplitudes of the Fourier harmonics R_n^2 is followed. One deals with a time series y_i of N observations. The null hypothesis is taken as the hypothesis that $E(y_i) = 0$ and that the y_i are independently and normally distributed with constant variance. Let the index n of R_n^2 denote the index of the fitted harmonic such that N/n denotes the period of this harmonic. If N/n is an integer one can replace the series of N terms by one of N/n *means* $\sum y_i/n$ where the y_i are of the same phase in period N/n. The harmonics of this series are selected harmonics of the original series. This paper examines the implications of such a breakdown for the testing of the significance of the short period harmonics relative to the null hypothesis.

5. Univariate Two-Population Distribution-Free Discrimination, David S. Stoller, RAND Corporation.

A univariate random variable, z, is defined by the composite cumulative distribution function, $F(z) = \theta F_1(z) + (1 - \theta)F_2(z); 0 \leq \theta \leq 1$. Restrict F_1 and F_2 to be such that the optimum a priori discriminating regions are $S_1 = \{z \mid z \leq \zeta\}$ and $S_2 = \{z \mid z > \zeta\}$, where ζ is unique. Optimum discrimination is defined as that which maximizes the probability of correctly classifying z. Denote the above maximum probability by $Q(\zeta)$. Given an independent random sample of size N from $F(z)$, each member of which is classifiable, a distribution-free estimate of ζ, denoted by ζ^*, is constructed as follows. Let $t(z) = k(z) - h(z)$, where $k(z)$ is the number of observations from the first population (i.e., that defined by F_1) which are less than or equal to z, and $h(z)$ is similarly defined for the second population. Then ζ^* is any value of z that maximizes $t(z)$. The estimate, ζ^*, possesses two asymptotically optimum properties: (1) the probability of correct classification induced by using ζ^* instead of ζ converges in probability to $Q(\zeta)$, and (2) the quantity, $t(\zeta^*)$, may be used to construct an estimate of $Q(\zeta)$ which converges in probability to $Q(\zeta)$.

6. New Types of Easily Constructed Partially Balanced Incomplete Block Designs, John Mandel and Marvin Zelen, National Bureau of Standards.

In the planning of experiments in the physical sciences one is often confronted with natural limitations on the size of experimental blocks. Therefore, the use of incomplete blocks is becoming ever more widespread in this type of application. In this paper a type of partially balanced incomplete block design is introduced, the construction of which consists in replacing each treatment of a balanced design with a group of treatments which themselves form a balanced design. A large class of designs thus becomes at once available by combining Latin Squares with Youden Squares, or Youden Squares with Youden Squares. An important property of these designs is the possibility of two-way elimination of error (according to rows and columns). A general formula is given for the Least-squares estimation of corrected treatment effects. Because of the flexibility of the proposed designs, their ease of construction, and simplicity of analysis, they are well adapted to experiments in physical and chemical laboratories. Investigations are in progress to extend the results to designs formed from combinations of chain block and other partially balanced incomplete block designs.

7. The Stochastic Convergence of a Function of Sample Successive Differences, Lionel Weiss, University of Virginia.

Let $f(x)$ be a bounded density function with at most a finite number of discontinuities, and such that there are two finite numbers, A and B $(A < B)$, with $f(x)$ nondecreasing in the interval $(-\infty, A)$ and nonincreasing in the interval (B, ∞). Let $X_1, X_2, \cdots, X_n$ be independent chance variables each with the density $f(x)$. Define $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ as the ordered values of $X_1, X_2, \cdots, X_n$; T_i as $Y_{i+1} - Y_i$; and $R_n(t)$ as the proportion of the values $T_1, \cdots, T_{n-1}$ not greater than $t/(n-1)$. $S(t)$ denotes $\left[1 - \int_{-\infty}^{\infty} f(x)e^{-tf(x)}\,dx\right]$; and $V(n)$ denotes $\sup_{t \geq 0} | R_n(t) - S(t) |$. Then it is shown that $V(n)$ converges stochastically to zero as n increases. This result can be used to demonstrate the stochastic convergence of various functions of $T_1, \cdots, T_{n-1}$, including some which have been discussed in the literature.

8. On a Modified T² Problem, Ingram Olkin, Michigan State College and S. S. Shrikhande, College of Science, Nagpur.

Consider two independent random vectors $X = (X_1, \cdots, X_p)$, $Y = (Y_1, \cdots, Y_p)$, each obeying a p-variate normal probability law with $EX = (\theta_1, \cdots, \theta_k, \mu_{k+1}, \cdots, \mu_p)$, $EY = (\theta_1, \cdots, \theta_k, \nu_{k+1}, \cdots, \nu_p)$, and same covariance matrix Σ, with all the parameters unknown. On the basis of a sample of n and m observations from X and Y, respectively, the hypothesis $H_0 : \mu_i = \nu_i$ against $H_1 : \mu_i \neq \nu_i$ $(i = k + 1, \cdots, p)$ is to be tested. The problem is equivalent to the case where X and Y are random vectors with means $EX = (0, \cdots, 0, \phi_{k+1}, \cdots, \phi_p)$, $EY = (0, \cdots, 0)$, and same covariance matrix Σ. On the basis of one and n observations from X and Y, respectively, $H_0 : \phi_i = 0$ against $H_1 : \phi_i \neq 0$ $(i = k + 1, \cdots, p)$ is to be tested. The likelihood ratio statistic is obtained and its distribution under H_0 and H_1 derived. If $k = 0$, the statistic reduces to Hotelling's T^2 statistic.

9. The Validity of Sheppard's Corrections for Grouping, F. J. Anscombe, University of Cambridge and Princeton University.

The moments of an absolutely continuous one-dimensional distribution are to be compared with the moments of the same distribution when it has been "grouped" with constant grouping interval. The characteristic function $\theta^*(t)$ of the grouped distribution may be expressed as a Fourier series in terms of the characteristic function $\theta(t)$ of the original distribution. The expansion is similar to those for moments given by R. A. Fisher (*Philos. Trans. Roy. Soc. London, Ser. A.*, Vol. 222 (1922), pp. 309–368, section 5), but requires no condition on the original distribution other than absolute continuity of the distribution function $F(x)$. Sheppard's formulas are obtained when the periodic terms in the series are neglected. The periodic terms are small if $| \theta(t) |$ is small for large t, and this condition is related to the differentiability of $F(x)$ for all values of x. The emphasis that has often been placed on the differentiability of $F(x)$ at infinity or at the ends of a finite range is misleading, because these points are not specially important.

10. Unbiased Tests Based on Unbiased Estimators, Reed B. Dawson, Department of Defense.

A test of a point-hypothesis $\theta = \theta_0$ of a distribution parameter will be said to be *strongly unbiased* when the power depends on θ alone and exceeds the size against all alternatives. For any α, $0 < \alpha < 1$, there exists a strongly unbiased test of size α if and only if there exists a real-valued function $f(\theta)$ which is zero at θ_0, strictly positive elsewhere, and pos-

sesses a bounded unbiased estimator. For, if $\omega(x)$ is the rejection probability corresponding to an outcome x, let $f = E\omega - \alpha$; if f is given, take $\omega = \alpha + K\hat{f}$, where $\hat{f}(x)$ is the estimator and K is a suitable positive constant. One application concerns a sample of n items from the family of all distributions over the unit interval. The possible strongly unbiased tests of a point-hypothesis on the rth moment form a bounded convex body in E_n over which the power is a linear functional. A second application (Mosteller's suggestion) concerns the hypothesis of independence of two attributes in a 2 x 2 table where sampling proceeds until a chosen cell attains a fixed quota. Powers of the determinant of the underlying probabilities admit bounded unbiased estimation, giving unbiased tests without the Neyman structure.

11. **The Mean Square Error of the Sample Median,** Harold Hotelling, University of North Carolina.

For random samples of any odd number from an arbitrary population, the ratio of the mean square error in the sample median, regarded as an estimate of the population median, to the corresponding population parameter, is shown never to be less than unity. This lower bound is actually attained for the familiar two-point distribution with equal probabilities. The fact that in this case the accuracy, however measured, of the median of a large number of observations is no better than that of one random observation destroys the argument sometimes given that the median should be used in the absence of knowledge of the form of the underlying distribution. (Research sponsored by the Office of Naval Research at Chapel Hill, North Carolina).

12. **The Moments of the Sample Median,** J. T. Chu and Harold Hotelling, University of North Carolina.

Moments of medians of random samples are studied by a method involving expansion about $\frac{1}{2}$ of the inverse of the cumulative distribution function, and in other ways. Readily calculable approximations are found, both for large and for small samples, with close upper and lower bounds on the errors of approximation. The asymptotic behvaior for large samples is examined. Calculations are carried out for the Laplace, Cauchy and normal distributions. (Research sponsored by the Office of Naval Research at Chapel Hill, North Carolina).

13. **Distribution of the Largest Vote in Unstructured Random Balloting,** Leo Katz, Michigan State College

The exact distributions of the maximum vote are obtained for two balloting arrangements. In both, each person votes once at random without prior reduction of the field of choice by a nomination process. In the first arrangement, a person may, if he (randomly) wishes, vote for himself; in the second, voting is gentlemanly. The second case has direct application to determination of "stars" in sociometric testing. An approximation is given; it is shown to be reasonably accurate for moderate-sized groups.

14. **Statistical Programming,** D. F. Votaw, Jr., Yale University.

Statistical programming problems arise when some of the constants in a programming problem are unknown but statistical information about them is available. In this paper several methods of statistical programming are compared in connection with a special linear programming problem. The application of simultaneous confidence interval estimation is discussed. (Work sponsored by the Office of Naval Research.)

15. Exact Tests of Significance for Combining Inter- and Intra-Block Information in Incomplete Block Designs (Preliminary Report), Marvin Zelen, National Bureau of Standards.

Consider an incomplete block design where the number of blocks is greater than the number of treatments ($b > v$). It is then shown under the usual assumptions for the recovery of inter-block information that two independent F tests of the null hypothesis (all treatments are the same) exist; one using only inter-block information and the other using the intra-block information. Let $F_i(i = 1, 2)$ represent the F ratio obtained for each test; $1 + r_i\lambda/\sigma_i^2$ represent the expected value of the numerator to the denominator of respective F ratios, where $\lambda = \Sigma(t_i - \bar{t})^2/v - 1$ is a measure of the departure from the null hypothesis (i.e., $\lambda = 0$ if H_0 is true); also let $p_i = P\{F \geqq F_i \mid H_0\}$. Then a combined test which seems to adjust for the differences in power of the two independent tests is given by the region $\{p_1p_2^\theta \leqq Q\}$, where Q is chosen such that $P\{p_1p_2^\theta \leqq Q\} = Q - \theta^{1/\theta}/1 - \theta = \alpha$ (level of significance), and $\theta = r_2\sigma_1^2/r_1\sigma_2^2$. For example, $\theta = 1 - E/E[\sigma^2/\sigma^2 + k\sigma_b^2]$ for balanced incomplete block designs where σ^2 is the "within block" variance, σ_b^2 the "between blocks" variance component, E is the efficiency factor and k is the plot size. Approximations to the power function of the test have been derived and preliminary calculations indicate that the above critical region seems to have greater power as compared to weighting the individual p_i's equally as in Fisher's method.

16. Moments and Related Quantities of the Null Distribution of Linear Contrasts of Order Statistics in the Case of Three Populations, Jacques St. Pierre, University of North Carolina.

Consider three independent normal populations with unknown means, m_0, m_1, m_2 respectively, and with a common known variance $\sigma^2 = 1$ say. Suppose a sample of size N is available from each population. Let $x_{(0)} > x_{(1)} > x_{(2)}$ be the ordered sample means. Consider the linear contrasts $z = x_{(0)} - cx_{(1)} - (1 - c)x_{(2)}$, where $0 \leqq c \leqq 1$. An expression for the kth moment about the origin is obtained. Properties of the moments and related quantities (skewness and kurtosis coefficients) are established, considering these quantities as functions of the nonstochastic parameter c. Tables of moments of low order are given in cases of special interest.

17. Application of Faà di Bruno's Formula in Mathematical Statistics, Eugene Lukacs, Office of Naval Research.

Let $z = G(y)$ and $y = f(x)$ be two functions such that all the derivatives of $G(y)$ and $f(x)$ up to order p exist. We denote by $D_t^k\{\ \}$ the operation of determining the kth derivative of the function in the braces with respect to t and we write $f_\nu = D_t^\nu\{f(t)\}/\nu!$, $f = f_0 = f(t)$ then $D_t^pz = D_t^p\{G[f(t)]\} = \Sigma p!D_y^k\{G(y)\}f_{k_1}^{i_1} \cdots f_{k_s}^{i_s}/(i_1! \cdots i_s!)$ where the summation is to be extended over all partitions of p such that $i_1 + i_2 + \cdots + i_s = k$ and $i_1k_1 + i_2k_2 + \cdots + i_sk_s = p$. This formula is due to F. Faà di Bruno [Sullo Sviluppo delli Funzioni, *Annali di science mathematiche e fisiche 6* (1853), pp. 479–480,]. Faà di Bruno's formula can be applied in mathematical statistics. The relations between the cumulants and the moments of a distribution are derived easily by means of this formula. It is also useful in the study of R. A. Fisher's k-statistics. For instance, the explicit formula, expressing the k-statistic of order p in terms of the observations, can be obtained. In addition to these familiar results, the following theorem is proven. Let $x_1, x_2, \cdots, x_n$ be n independent observations taken from a population with distribution function $F(x)$ and denote by p an integer greater than one. Assume that the pth moment of $F(x)$ exists. The population is normal if, and only if, the k-statistic of order p is independent of the sample mean.

18. On Simultaneous Minimax Point Estimation, Waldo A. Vezeau and Koichi Ito, St. Louis University.

This paper is concerned with simultaneous minimax point estimation of all the parameters in the multivariate distribution function of a parent population on the basis of a sample of fixed size. Extending results due to K. Miyasawa (*Bull. Math. Stat.*, Vol. 5(1953), pp. 1–17), it is shown that if the risk is a bounded function of s parameters, $\theta_1, \cdots, \theta_s$ and their point estimates, $d_1, \cdots, d_s$, and a convex, measurable function of $d_1, \cdots, d_s$ for any fixed $\theta_1, \cdots, \theta_s$, and if the space D of $d_1, \cdots, d_s$ is compact and convex, then there exists a set of simultaneous minimax point estimates of $\theta_1, \cdots, \theta_s$ in D. Applications of this theorem are made to simultaneous minimax point estimation of the parameters in a multinomial distribution, the mean and variance (or standard deviation) of a univariate normal distribution, and the means, variances and covariances of a multivariate normal distribution.

19. Estimation of Structural Parameters when the Number of Incidental Parameters is Unbounded, J. Wolfowitz, Cornell University.

Let $\prod_{i=1}^{n} \prod_{j=1}^{m_i} f(z_{ij} \mid \theta, \alpha_i)$ be the frequency function of the observed chance variables $\{X_{ij}\}$, $i = 1, \cdots, n$; $j = 1, \cdots, m_i$, which depends upon the unknown (structural) parameter θ and the unknown (incidental) parameters $\{\alpha_i\}$. The author proves that in general there exists no estimator of θ which is efficient for all sequences $\{\alpha_i\}$. This verifies a conjecture of the author's, described in the *Proc. Roy. Dutch Acad. Sci.*, Ser. A, Vol. 56, No. 2, and *Indag. Math.*, Vol. 15, 1953, where a heuristic supporting argument was given.

20. On Power Properties of Certain Simultaneous Tests, K. V. Ramachandran, University of North Carolina.

(1) Let $y_1, y_2, \cdots, y_K$ be k independent normal variates with $E(y_i) = \mu_i$ and $v(y_i) = \sigma^2 (i = 1, 2, \cdots, K)$. μ_i and σ^2 are unknown but an independent estimate s^2 of σ^2 with v d.f. is available. To test the hypothesis: $\mu_1 = \mu_2 = \cdots = \mu_K$ we have a short cut test of Tukey based on the studentized range. (2) Let $y_1, y_2, \cdots, y_K$ be k independent normal variates with variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_K^2$ respectively. To test the hypothesis: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$ we have the F_{max} ratio test of Hartley. In this paper the following properties of the tests are proved. The power function of the tests depend only on $k - 1$ parameters, namely, $\delta_{i-1} = \mu_i - \mu_1 (i = 2, 3, \cdots, K)$ in case (1) and $\eta_{i-1} = \sigma_i^2/\sigma_1^2$ $(i = 2, 3, \cdots, K)$ in case (2). The tests are completely unbiased but the power functions do not have the monotonicity property. A set of useful lower bounds are obtained for the power in the two situations. Power properties of multivariate and other generalizations of these tests are being investigated.

21. On Tests of Normality and Other Tests of Goodness of Fit Based on Distance Methods, M. Kac, J. Kiefer, and J. Wolfowitz, Cornell University.

The authors study the problem of testing whether the common distribution function (d.f.) of the observed independent chance variables $x_1, \cdots, x_n$ is a member of a given class. A classical problem is concerned with the case where this class is the class of all normal d.f.'s, and for the sake of brevity the description in this abstract will be limited to some of the results for this problem. For any two d.f.'s $F(y)$ and $G(y)$ let $\delta(F, G) = \sup_y \mid F(y) - G(y) \mid$. Let $N(y \mid \mu, \sigma^2)$ be the normal d.f. with mean μ and variance σ^2. Define $\bar{x} = n^{-1}\sum_1^n x_i$, $s^2 = n^{-1}\sum_1^n x_i^2 - \bar{x}^2$. Let $G_n^*(y)$ be the empiric d.f. of $x_1, \cdots, x_n$. The authors consider, inter alia, tests of normality based on $v_n = \delta(G_n^*(y), N(y \mid \bar{x}, s^2))$ and on $w_n = \int (G_n^*(y) - N(y \mid \bar{x}, s^2))^2 \, d_y N(y \mid \bar{x}, s^2)$. It is shown that the asymptotic power

of these tests is considerably greater than that of the optimum χ^2 test. The covariance function of a certain Gaussian process $Z(t)$, $0 \leq t \leq 1$, is found. It is shown that the sample functions of $Z(t)$ are continuous with probability one, and that, as $n \to \infty$, $\lim P\{nw_n < a\} = P\{W < a\}$, where $W = \int_0^1 [Z(t)]^2\, dt$. Tables of the distribution of W and of the limiting distribution of $\sqrt{n}\, v_n$ are given. The role of various metrics is discussed.

22. **Tolerance Regions (Preliminary Report),** D. A. S. Fraser and I. Guttman, University of Toronto.

Three definitions are considered for tolerance regions. A "distribution-free tolerance region" has the distribution of its probability content independent of the parameter. A "β-content tolerance region" has probability content at least β with an assigned level of confidence. A "β-expectation tolerance region" has probability content on the average equal to β. For the first definition a necessary and sufficient condition has been obtained for the characteristic function of the region. For sampling from univariate distributions for which the order statistics are complete, the nonexistence of distribution-free tolerance regions was obtained in the discontinuous case and some results on distribution-free tolerance bounds were obtained in the continuous case. For the third definition an analogy with hypothesis testing has been established by introducing a density function to indicate the desirability that different points of a distribution be included in the region. For normal distributions the center of the distribution was weighted more heavily than the tails and most stringent tolerance regions obtained. For univariate distributions they were $[\bar{X} \pm \lambda\sigma]$ and $[\bar{X} \pm \lambda s_x]$ depending on whether or not σ was known. In the multivariate case they are based on Hotelling's T^2 statistic.

23. **Comparison of the Power of Nonparametric Two Sample Tests against Normal Alternatives,** Benjamin Epstein, Wayne University.

This is a sampling study in which we compare the power of run, rank sum, exceedance, and truncated maximum deviation two sample tests. The particular case studied involves normal alternatives whose distance apart is measured by the difference in population means. Two hundred random samples of size 10 are drawn from each population. These results are related to recent work of Dixon and Teichroew [Abstract, *Ann. Math. Stat.* Vol. 25, (1954), p. 175]. There are, however, these differences: (i) in the present study we assume that the two samples have been placed (simultaneously) on life test, thus making the times to failure available in an ordered way and (ii) include exceedance and truncated maximum deviation rules among the nonparametric tests. Such rules are particularly useful in life test situations. Experimental sampling assigns the following order to the power (best to worse): rank sum, untruncated maximum deviation, truncated maximum deviation, exceedance, and run. The first four power curves are fairly close together and are all substantially better than the power curve for the run test. Also included in the paper is experimental information on the expected number of items failed in reaching a decision when an exceedance or truncated maximum deviation rule is used. Substantial savings in this direction are possible. (Research sponsored by the Office of Ordnance Research, U. S. Army).

24. **On the Distribution of Radial Errors Having Normally Distributed Components,** A. C. Cohen, Jr., University of Georgia.

For a set of p independent random variables x_j $(j = 1, 2, \cdots, p)$, each of which is normal $(0, \sigma)$, the radial error defined as $r = (x_1^2 + x_2^2 + \cdots + x_p^2)^{1/2}$ is considered. It is well known that the distribution of r is given by $[2r/\sigma^2]f_p(r^2/\sigma^2)$ where $f_p(\chi^2)$ is the χ^2 frequency function with p degrees of freedom. This paper is concerned with the problem of estimating the scale parameter σ from unrestricted (complete), truncated, and censored samples

of r. Maximum likelihood estimators are developed for each of these cases, and asymptotic estimate variances are given. In the case of unrestricted samples, $(pn\hat{\sigma}^2/\sigma^2)$ has a χ^2 distribution with pn degrees of freedom, where n is the number of sample observations and $\hat{\sigma}^2$ is the maximum likelihood estimate. Tables and graphs of functions necessary for solving the maximum likelihood estimating equations for truncated samples are given for $p = 2$ and $p = 3$. Illustrative examples relating to target analysis studies are included.

25. **Confidence Bounds on Departures from a Particular Kind of Multi-Collinearity of Means,** S. N. Roy, University of North Carolina.

For k $(p + q)$-variate $N(\xi_i, \Sigma)$, where $\Sigma((p + q) \times (p + q))$ is symmetric p.d. with submatrices $\Sigma_{11}(p \times p)$, $\Sigma_{22}(q \times q)$ and $\Sigma_{12}(p \times q)$, and $\xi_i((p + q) \times 1)$ has column subvectors $\xi_{1i}(p \times 1)$ and $\xi_{2i}(q \times 1)$, we can set, in the following way, confidence bounds on $\xi_{1i} - \Sigma_{12}\Sigma_{22}^{-1}\xi_{2i}$ which are departures from the hypothesis $\xi_{1i} - \Sigma_{12}\Sigma_{22}^{-1}\xi_{2i} = 0$ $(i = 1, 2, \cdots, k)$. Let S_{11}, S_{22} and S_{12} stand for the submatrices of the "within" covariance matrix pooled from k samples of size n each and $\underline{x}_{1i}(p \times 1)$ and $\underline{x}_{2i}(q \times 1)$ $(i = 1, \cdots, k)$ for the subvectors of the k sample mean vectors. Then setting $S_{1.2} = S_{11} - S_{12}S_{22}^{-1}S_{12}'$, and $B(p \times k)$ and $\beta(p \times k)$ for the matrices with respective column vectors $\underline{x}_{1i} - S_{12}S_{22}^{-1}\underline{x}_{2i}$ and $\xi_{1i} - \Sigma_{12}\Sigma_{22}^{-1}\xi_{2i}$ $(i = 1, \cdots, k)$, we have, with a confidence coefficient, say $1 - \alpha$, the following set of simultaneous confidence bounds (for all arbitrary nonnull $\underline{a}'(1 \times p)$ and unit-length $\underline{b}(k \times 1)$): $\underline{a}'B\underline{b} - [k(\underline{a}'S_{1.2}\underline{a})c_\alpha(p, k, nk - k)]^{1/2} \leq \underline{a}'\beta\underline{b} \leq \underline{a}'B\underline{b} + [k(\underline{a}'S_{1.2}\underline{a})c_\alpha(p, k, nk - k)]^{1/2}$, where $c_\alpha(p, k, nk - k)$ is the upper α-point of the distribution of the (central) largest determinantal root based on p, k and $nk - k$ D. F. Test for the associated hypothesis is also easily obtained.

26. **The Efficiency of Tests,** Wassily Hoeffding and Joan R. Rosenblatt, University of North Carolina.

The efficiency of a family of tests is defined. Let $\{X_n\}$ be a sequence of random variables such that for every n the vector $(X_1, \cdots, X_n)$ has cdf G_n in some class $\mathcal{C}_n$. Let $\mathcal{C}_{1n}$, $\mathcal{C}_{2n}$ be disjoint subsets of $\mathcal{C}_n$ such that we prefer one or the other of two alternatives A_1, A_2 according as $G_n \epsilon \mathcal{C}_{in}$ $(i = 1, 2)$. Given α_1, α_2, we say that the problem $(\{\mathcal{C}_{1n}\}, \{\mathcal{C}_{2n}\}, \alpha_1, \alpha_2)$ is solved by a test (general nonsequential two-decision rule) ϕ_n such that $P(\phi_n \text{ selects } A_i \mid G_n) \geq 1 - \alpha_i$ for all $G_n \epsilon \mathcal{C}_{in}$ $(i = 1, 2)$. The index of efficiency of a family of tests $\mathfrak{J}$ for the problem $(\{\mathcal{C}_{1n}\}, \{\mathcal{C}_{2n}\}, \alpha_1, \alpha_2)$ is $N(\mathfrak{J}) = N(\mathfrak{J}, \{\mathcal{C}_{1n}\}, \{\mathcal{C}_{2n}\}, \alpha_1, \alpha_2)$, the least sample size with which the problem can be solved by a member of the family $\mathfrak{J}$. If $\mathfrak{J}_1$, $\mathfrak{J}_2$ are two families of tests, the efficiency of $\mathfrak{J}_2$ relative to $\mathfrak{J}_1$ is given by eff $(\mathfrak{J}_2/\mathfrak{J}_1) = N(\mathfrak{J}_1)/N(\mathfrak{J}_2)$. The determination of $N(\mathfrak{J})$ is closely related to finding a test which maximizes the minimum power. Let $\theta(G_n)$ be a real-valued function of G_n and suppose $\mathcal{C}_{1n} = \{G_n : \theta(G_n) \leq \theta_1\}$, $\mathcal{C}_{2n} = \{G_n : \theta(G_n) \geq \theta_2\}$, $\theta_1 < \theta_2$. Under suitable assumptions, we derive asymptotic expressions for $N(\mathfrak{J})$ as $\delta = \theta_2 - \theta_1$ tends to zero while α_1, α_2 remain fixed.

27. **On a Decision Procedure to Select the Population with the Largest Mean (Preliminary Report),** R. C. Bose and Jacques St. Pierre, University of North Carolina.

Consider $n + 1$ independent normal populations with unknown means $m_0 \geq m_1 \geq m_2 \cdots \geq m_n$, respectively, and with known or unknown common variance σ^2. Suppose a sample of size N is available from each population, and a decision procedure is required to select the population with the largest mean, with the following properties. (a) Either a decision is made that the population from which the ith sample was drawn has the largest mean, or no decision is made. (b) The probability of making a wrong decision (if a decision is made) is less than a pre-assigned number α_0 (independent of the unknown means

$m_0, m_1, \cdots, m_n$. Subject to the requirements (a) and (b), the decision rule must control the chance of indecision. The case of three populations with known σ^2 is considered in detail, and the properties of a decision rule based on the auxiliary statistic $y = x_{(0)} - x_{(1)}$ are studied, where $x_{(0)} \geqq x_{(1)} \geqq x_{(2)}$ are the ordered sample means, the rule being to decide that $x_{(0)}$ comes from the population with the largest mean if $y > k$, and not to take a decision if $y \leqq k$. The general case when $n > 2$ and σ^2 is unknown is under consideration.

28. **Most Economical Multiple-Decision Rules,** William Jackson Hall, University of North Carolina.

Suppose x has an unknown distribution function F, belonging to one of m disjoint classes $\omega_1, \cdots, \omega_m$, and suppose $A_1, \cdots, A_m$ are corresponding alternative decisions. A decision rule D_N, based on a sample of size N, is said to be a "most economical multiple-decision rule (M.E. d.r.) relative to $(\alpha_1, \cdots, \alpha_m)$, $0 \leqq \alpha_i < 1$, for choosing among $A_1, \cdots, A_m$" if it satisfies (1) $\Pr(D_n \text{ chooses } A_i \mid F) \geqq \alpha_i$ for all $F \in \omega_i (i = 1, \cdots, m)$ and if N is the least integer n for which (1) can be satisfied. It is proved that to obtain M.E. d.r.'s one need only consider d.r.'s in the sequence $\{D_n^0\}$, $n = 0, 1, 2, \cdots$, where D_n^0 denotes a minimax solution w.r.t. a certain weight function for samples of fixed size n. If ω_i contains but one distribution F_i $(i = 1, \cdots, m)$, D_n^0 is of the form: (2) choose A_i if $a_iL_i \geqq a_jL_j$ $(j = 1, \cdots, m)$ where $L_1, \cdots, L_m$ are the likelihood functions of the sample corresponding to $F_1, \cdots, F_m$ and $a_1, \cdots, a_m$ are positive constants. In the general case, D_n^0 is of a similar form where now $F_1, \cdots, F_m$ are "average" distribution functions, averaged w.r.t. least favorable conditional distributions over $\omega_1, \cdots, \omega_m$ (if existent). Similar results are obtained for "M.E. d.r.'s relative to (β_{ij}), $0 < \beta_{ij} \leqq 1$," defined as above with (1) replaced by (1') $\Pr(D_n \text{ chooses } A_i \mid F) \leqq \beta_{ij}$ for all $F \in \omega_j (i \neq j; j = 1, \cdots, m)$; and (2) is replaced by: (2') choose A_i if $\sum_{k \neq i} b_kL_k \leqq \sum_{k \neq j} b_kL_k$ $(j = 1, \cdots, m)$, for some positive constants $b_1, \cdots, b_m$. Other properties of the d.r.'s are derived and various extensions and examples given.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Archie Blake is now employed as an Advisory Engineer in the Systems Analysis of Westinghouse Electric Corporation, Baltimore, Maryland.

E. L. Cox has left Operations Research Group, Case Institute of Technology, to take a position with Chemical Corps Biological Laboratories, Frederick, Maryland.

Harold Davis has transferred from Headquarters, United States Air Force to The Operations Analysis Office, Hq. Far East Air Forces.

Professor Hilda Geiringer is on leave of absence from Wheaton College in order to complete and prepare for publication on behalf of Harvard University some of the post-humous work of Richard von Mises.

Dr. S. G. Ghurye has accepted the position of Reader in Statistics, Department of Mathematics and Statistics, University of Lucknow, Lucknow, U.P., India.

The University of London has conferred the D.Sc. degree in Mathematical statistics on H. O. Hartley for his published contributions in the field of mathematical statistics, and he has accepted a position as professor on the permanent staff of the Department of Statistics and Statistical Laboratory, Iowa State College. Dr. Hartley has been visiting professor of statistics at Iowa State College since July 1, 1954.

Stanley L. Isaacson has returned to his position as Assistant Professor of Statistics at Iowa State College after spending a year on leave of absence as Visiting Assistant Professor of Statistics at Stanford University.

Dr. Hans Kellerer was nominated as fulship Professor for Statistics and as Direktor of the Statistical Seminar at the Wirtschafts- und Sozialwissenschaftliche Fakultät of the Freie Universität in Western-Berlin 1st of April 1953.

Dr. Nathan Keyfitz has resumed his duties as Senior Research Statistician of the Dominion Bureau of Statistics after a year's assignment with the Government of Indonesia.

Professor T. C. Koopmans will be on leave of absence from the University of Chicago and the Cowles Commission during the academic year 1954–55. He will spend this year at Yale University in teaching and research.

Dr. Richard F. Link has left the Analytical Research Group at the Forrestal Research Center, Princeton University to accept a position with the Sandia Corporation of Albuquerque, New Mexico.

Lt. Carl R. Ohman, formerly a graduate student at Princeton University, is now on active duty with the U. S. Army, stationed in Washington, D. C.

Donald B. Owen has resigned as Assistant Professor of Mathematics at Purdue University to accept a position as a Staff Member with the Sandia Corporation in Albuquerque, New Mexico.

K. C. S. Pillai, who was a Research Associate in the Department of Statistics, University of North Carolina, joined the Statistical Office of United Nations, New York in February, 1954. He has completed his work for Ph.D. degree in Statistics in the University of North Carolina.

John W. Richardson is employed as a Physicist at the Ramo-Wooldridge Corporation, Los Angeles, California.

Daniel E. Sands has accepted a position as Biometrician in the Statistics Section of the Squibb Institute for Medical Research, New Brunswick, New Jersey.

Major Oliver A. Shaw is now stationed at Hq. Air Research and Development Command where he is serving as Research Administrator in Mathematics and Mathematical Statistics.

George W. Snedecor has been elected an Honorary Fellow of the Royal Statistical Society "in consideration of the eminent services rendered to statistics."

James H. Straughan of Michigan State College has accepted a position as Assistant Professor, Department of Psychology, Montana State University, Missoula, Montana.

Dr. Joseph V. Talacko, Assistant Professor of Mathematics, is on leave of absence from the Marquette University, Milwaukee for the academic year 1954–55. He has a Ford Foundation Fellowship and plans to spend most of the year at the Statistical Laboratory, University of California in Berkeley and to visit several west coast universities.

Dr. Fred H. Tingey, formerly with the General Electric Company, Richland, Washington, has joined the staff of the Atomic Energy Division of Phillips Petroleum Company at Idaho Falls, Idaho where he will be responsible for all statistical studies arising from plant operations.

M. C. Throdahl, formerly Development Manager for Rubber Chemicals, Monsanto Chemical Company at Nitro, West Virginia, has been given a new assignment as Assistant Director of the company's Development Department in St. Louis.

Elizabeth Vaughan has transferred from the U. S. Fish and Wildlife Service to the Quality Evaluation Laboratory, Naval Ammunition Depot, Bangor, Washington, as an Analytical Statistician and head of the section on special investigations.

David L. Wallace, formerly at Massachusetts Institute of Technology, has been appointed an Assistant Professor of Statistics at the University of Chicago.

Samuel Weiss, formerly Chief Statistician and Chief of the Office of Statistical Standards, Bureau of Labor Statistics, has recently established a private statistical consulting office in Washington, D. C. He will, however, continue to act as a consultant to the Commissioner of Labor Statistics.

A. W. Wortham has completed the requirements for a Ph.D. degree at Oklahoma A. and M. College and has returned to his position with Chance Vought Aircraft, Dallas, Texas.

R. I. Piper of the Pacific Telephone and Telegraph Company, San Francisco, California, died on October 20, 1953 at the age of fifty-one years. He was a member of the Institute for ten years.

Fellowships in Statistics, University of Chicago

Members of the Institute of Mathematical Statistics are invited to nominate research workers whom they feel could benefit from three $4000 post-doctoral fellowships in Statistics offered for 1955–56 by the University of Chicago. The purpose of these fellowships, which are open to holders of the doctor's degree or its equivalent in research accomplishment, is to acquaint established research workers in the biological, physical, and social sciences with the role of modern statistical analysis in the planning of experiments and other investigative programs, and in the analysis of empirical data. The development of the field of Statistics has been so rapid that most current research falls far short of attainable standards, and these fellowships (which represent the fifth year of a five-year program supported by The Rockefeller Foundation) are intended to help reduce this lag by giving statistical training to scientists whose primary interests are in substantive fields rather than in Statistics itself. Nominations, which should be

made soon since the closing date for applications is February 15, 1955, may be addressed to any member of the University of Chicago Statistics Department, or to its Chairman, W. Allen Wallis.

First National Meeting, Society for Industrial and Applied Mathematics

The Society for Industrial and Applied Mathematics will hold its first national meeting in conjunction with the annual meetings of the American Mathematical Society, the Mathematical Association of America, and the Association for Symbolic Logic at the University of Pittsburgh, on December 27–29. The following addresses will be presented to an evening meeting: "The History of a Problem" Dr. Brockway McMillan, Bell Telephone Laboratories; "The Control of Industrial Operations", Professor Herbert A. Simon, Carnegie Institute of Technology; "Probability Theory in Liability and Property Insurance", Mr. C. W. Crouse, Actuary, Preslan and Company. Further information can be obtained from H. W. Kuhn, Dalton Hall, Bryn Mawr College, Bryn Mawr, Pennsylvania.

New Members

The following persons have been elected to membership in the Institute

May 14, 1954 to August 9, 1954

Abbott, James H., M.S. (Southern Methodist Univ.), Graduate Student, University of Illinois, *Box 64, University Station, Urbana, Illinois.*

Atkinson, Richard C., Ph. B. (Univ. of Chicago), Research Assistant, Psychology Department, Indiana University, Bloomington, Indiana.

Austin, Thomas L., Jr., BBA (Univ. of Georgia), Graduate Assistant, University of Georgia, College of Business Administration, Athens, Georgia, *195 Thirteenth Street, N.E., Atlanta, Georgia.*

Baade, William H., B.S. (Mass. College of Pharmacy), Chemist, Process Research and Development, Merck and Company, Inc., Rahway, New Jersey.

Barnes, Gerald W., M.A. (Univ. of Arkansas), Teaching Fellow, Indiana University, Department of Psychology, Bloomington, Indiana.

Binford, J. R., A.B. (DePauw), Teaching Fellow, Indiana University, Psychology Department, Bloomington, Indiana.

Blumenthal, Robert M., B.A. (Oberlin College), Graduate Student and Teaching Assistant, Department of Mathematics, White Hall, Cornell University, Ithaca, New York.

Box, George E. P., Ph.D. (London), Statistician, Statistical Research Section, M.C.S.D., Imperial Chemical Industries, Dyestuffs Division, Blackley, Manchester 9, England.

Bramwell, W. K., Jr., B.S. (Arizona Univ.), Vice President, Hardin County Savings Bank, Eldora, Iowa.

Brunelle, Robert H., B.A. (Univ. of the State of New York, Champlain College), Graduate Student, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Buchbinder, Benjamin, B.A. (Brooklyn College), Graduate Student, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Carterette, E. C. H., B.A. (Harvard Univ. Honors), Graduate Student and Research As-

sistant, Hearing and Communication Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana.

Chu, John T., Ph.D. (Iowa State College), Research Associate, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina, *Box 132, Chapel Hill, North Carolina.*

Cisin, Ira H., M.A. (American Univ.), Director of Research in Motivation, Morale and Leadership, Human Resources Research Office, George Washington University, 2013 G Street N.W., Washington, D. C.

Colton, Theodore, B.A. (Brooklyn College), Graduate Student, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Coons, Irma J., B.S. (Harding College), Graduate Student, Iowa State College, Ames, Iowa, *Bevier House, Iowa State College, Ames, Iowa.*

Davis, Don Allen, B.A. (Western Washington College of Education), Teaching Assistant, University of Washington, Seattle 5, Washington, *3748½ University Way, Seattle 5, Washington.*

Dillon, Thaddeus, III, M.S. (John Carroll Univ.), Instructor in Mathematics, John Carroll University, Cleveland 18, Ohio.

Doan, Elizabeth, A.B. (Oberlin College), Graduate Student in Mathematical Statistics, University of North Carolina, Chapel Hill, North Carolina, *Kenan Dormitory, Chapel Hill, North Carolina.*

Doto, Irene L., M.A. (Temple Univ.), Instructor, Temple University, Broad Street and Montgomery Avenue, Philadelphia 22, Pennsylvania, *5949 Nassau Road, Philadelphia 31, Pennsylvania.*

Ecimovic, Juraj P., diploma (Faculty of Economics, Zagreb), Expert on Statistical Sampling in Agriculture of the Food and Agriculture Organization of the United Nations, Technical Assistance Mission to Indonesia, Djakarta, Djalan, Hajam Wuruk 6, Indonesia.

Ferguson, Barbara June, B.S. (Univ. of Washington), Graduate Student, Mathematics Department, University of Washington, Seattle, Washington, *2736 60th Avenue, S.W., Seattle 6, Washington.*

Frankmann, Raymond W., A.B. (Harvard Univ.), Graduate Student, Department of Psychology, Indiana University, Bloomington, Indiana.

Free, Spencer M., Jr., Ph.D., (North Carolina State College), Research Statistician, Smith Kline & French Laboratories, 1530 Spring Garden Street, Philadelphia 1, Pennsylvania.

Gart, John J., B.S. (DePaul Univ.), Graduate Assistant, Mathematics Department, Marquette University, Milwaukee 3, Wisconsin, *4925 N. Hamilton Avenue, Chicago 25, Illinois.*

Goen, Richard L., B.S. (Univ. of Washington), Graduate Student and Teaching Fellow, University of Washington, Seattle 5, Washington, *4407 Densmore, Seattle 3, Washington.*

Hudson, John B., B.A. (Univ. of Oregon), Graduate Student and Research Assistant, Department of Sociology, University of Washington, Seattle 5, Washington.

Imhof, Jean Pierre, Licence ès Sciences Mathematiques (Univ. of Geneva, Switzerland), Graduate Student, University of California, Berkeley, California, *2310 Ellsworth Apt. 1, Berkeley 4, California.*

Iqbal, M., M.A. (Panjab Univ., Lahore, Pakistan), Senior Lecturer in Statistics, Panjab University, Lahore, Pakistan, *Statistics Department, University of North Carolina, Chapel Hill, North Carolina.*

Jerger, James F., M.A. (Northwestern Univ.), Research Audiologist, Northwestern University, Evanston, Illinois, Hearing Clinic, Speech Annex Building, Northwestern University, Evanston, Illinois.

Jones, Lawrence, B.S. (Denison Univ.), Graduate Student, Statistics Department, Iowa State College, Ames, Iowa, *2010 McCarthy Road, Ames, Iowa.*

Kelley, H. Paul, M.A. (Princeton Univ.), Psychometric Fellow, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

Krakowski, Martin, Ph. D. (Carnegie Inst. of Tech.), Research Mathematician, California Research Corporation, P. O. Box 446, LaHabra, California.

Krueger, William D., B.B.A. (Western Reserve Univ.), Market Analyst, Harris-Seybold Company, 4510 East 71st Street, Cleveland 5, Ohio.

Kudō, Hirokichi, Rigakushi (Tokyo Univ.), Assistant Professor of Mathematics, Ochanomizu University, Department of Mathematics, Ootsukamachi, Bunkyo-ku, Tokyo, Japan.

Lancaster, Henry Oliver, Ph.D. (Univ. of Sydney), Lecturer in Medical Statistics, School of Public Health and Tropical Medicine, University Post Office, Sydney, Australia.

Lehman, Eugene, H., Jr., M.A. (Teachers College Columbia Univ., Graduate Student in Mathematical Statistics and Research Assistant in Statistics, Highway Cost Allocation Project, Civil Engineering Department, University of Washington, Seattle 5, Washington, *3408 Goldendale Place, Union Bay Villege, University of Washington, Seattle 5, Washington.*

Lemus, Ferdinand, B.A. (Michigan State Normal College), Graduate Student in statistics at Iowa State College, Ames, Iowa.

McKean, Harlley E., M.S. (Purdue Univ.), Research Assistant, Purdue University, Lafayette, Indiana, *5-4 Ross-Ade Drive, West Lafayette, Indiana.*

Maggio, Ralph A., A.B. (Boston Univ.), Graduate Student, Applied Statistics, Rutgers University, New Brunswick, New Jersey, *Apartment 42-A, Codrington Apartments, Bound Brook, New Jersey.*

Meshadani, Mahmoud H., M.S. (Iowa State College), Graduate Student, Iowa State College, Ames, Iowa, *44d/57 Jami Atta, Baghdad, Iraq.*

Miller, Clark T., M.A. (Univ. of Wisconsin), Graduate Student, Department of Mathematics, University of Wisconsin, Madison 6, Wisconsin, *1553 Adams Street, Madison 5, Wisconsin.*

Moser, Joseph M., B.A. (St. John's Univ.), Graduate Student and Assistant Instructor, St. Louis University, St. Louis 3, Missouri, *359 N. Whittier Street, St. Louis 8, Missouri.*

Na Nagara, Prasert, M.S. (Cornell Univ.), Graduate Student, Cornell University, Ithaca, New York, *107 Harvard Place, Ithaca, New York.*

Noel, Lionel M., B.S. (U.S. Naval Academy), Lieutenant, U. S. Navy, Washington, D. C., *413-A Devereux Avenue, Princeton, New Jersey.*

Page, Ewan S., M.A.(Cantab.), Lecturer in Mathematical Statistics, *55 Stanley Drive, Leicester, England.*

Peterson, Henry H., B.S. (Univ. of Washington), Graduate Student and Teaching Assistant, University of Washington, Seattle, Washington, *8336 32nd Avenue, N.W., Seattle 7, Washington.*

Porter, Lee Burton, MBA (Univ. of Michigan), Actuarial Assistant, North Carolina Mutual Life Insurance Company, P. O. Box 201, Durham, North Carolina.

Richardson, Jack M.S. (Northwestern Univ.), Graduate Student, Psychology Department, Northwestern University, Evanston, Illinois.

Salveson, Melvin E., Ph.D. (Univ. of Chicago), Procedures Research Manager, Major Appliance Division, General Electric Company, Appliance Park, Louisville 1, Kentucky.

Schwartz, Lorraine, B.A. (Univ. of California), Research Assistant, Statistical Laboratory, University of California, Berkeley, California, *2208 Bancroft Way, Berkeley 4, California.*

Shapiro, Marvin, B.A. (Cornell Univ.), Graduate Student, Columbia University, New York 27, New York, *3202 Cleveland Avenue, Niagara Falls, New York.*

Sheldahl, Loren R., B.A. (Iowa State Teachers College), Instructor, School of Public Health, University of Minnesota, Minneapolis 14, Minnesota.

Soto-Alarcon, Jose L., B.S. (Univ. of Puerto Rico), Graduate Student, The Johns Hopkins University, 615 N. Wolfe St., Baltimore 5, Maryland.

Strehler, Allen F., Ph.D. (Univ. of Wisconsin), Head, Department of Mathematics, The Lovelace Foundation for Medical Education and Research, 4800 Gibson Boulevard, S.E., Albuquerque, New Mexico.

Takano, Kinsaku, Rigakushi (Faculty of Science, Osaka Univ.), Member of the Institute of Statistical Mathematics, *No. 203 Toei Apartments, 1548 Hatagaya-Nakamachi, Shibuya-ku, Tokyo, Japan.*

Tamargo, M. A., M.S. (Louisiana State Univ.), Head of the Grain & Root Crop Bureau (Acting Chief Statistics Department, Fiber Coop. Commission, Agricultural Experiment Station, Santiago Vegas, Cuba), Ministerio de Agricultura, Havana, Cuba, *Linea 656 altos ent. A y B, Vedado-Havana, Cuba.*

Taylor, Howard L., M.S. (Iowa State College), Associate and Student, Statistical Laboratory, Iowa State College, Ames, Iowa, *1323 Carroll Avenue, Ames, Iowa.*

Tucker, Thomas F., M.B.A. (Univ. of Pennsylvania, Wharton School), Assistant Actuary, Commonwealth of Pennsylvania, Insurance Department, 600 Packard Building, Philadelphia 2, Pennsylvania.

Van Der Byl, Willem, Doctor (State Univ. of Utrecht, Netherlands), Research Associate KNMI, Koninklyk Nederlands Meteorologisch Instituut, KNMI, DeBilt, Netherlands.

Van Natta, Pearl A. (Mrs. E.A.), M.A. (Univ. of Oregon), Research Assistant in Mathematical Statistics, University of Oregon, Eugene, Oregon.

Weber, Donald C., M.S. (Univ. of Wisconsin), Instructor in Mathematics, North Dakota Agricultural College, State College Station, Fargo, North Dakota, *414 S. Seventh Avenue, Wausau, Wisconsin.*

Weigel, E. Charles, B.S. (Rutgers Univ.), Accountant, Harry Rubenstein, C.P.A., 313 State Street, Perth Amboy, New Jersey, *423 Compton Avenue, Perth Amboy, New Jersey.*

Wertz, Frederick E., B.S. (Manhattan College), Graduate Student, Rutgers University, New Brunswick, New Jersey, *121 West 69th Street, New York 23, New York.*

White, Robert F., M.S. (Univ. of Connecticut), Graduate Assistant and Student, Iowa State College, Statistical Laboratory, Ames, Iowa.

Witeck, Eugene F., B.S. (Marquette Univ.), Graduate Student, University of Washington, Seattle, Washington, *8335 13th Avenue, N.W., Seattle 7, Washington.*

Young, Gregory O., M.S. (California Inst. of Tech.), Research Engineer, Technical Staff, Hughes Aircraft Company, Department 04, Culver City, California.

Zastrow, Marvin C., B.S. (Wisconsin State College), Graduate Assistant in mathematics, Marquette University, Milwaukee, Wisconsin, *1901 S. 24th Street, Milwaukee 4, Wisconsin.*

REPORT OF THE MONTREAL MEETING OF THE INSTITUTE

The sixteenth summer meeting, 62nd meeting of the Institute of Mathematical Statistics was held in Montreal, Canada, September 10–13, 1954. Other organizations meeting in Montreal at the same time were the American Statistical Association, the Econometric Society, the American Society for Quality Control (Montreal Section), and the Biometric Society (Eastern North American Region). On the evening of September 10, the City of Montreal held a reception for the members of the societies, and on the evening of September 11, there was an informal party for the members of the societies.

The following 199 members of the Institute registered for the meeting:

Helen Abbey, R. L. Anderson, F. J. Anscombe, K. J. Arnold, J. C. Bain, J. D. Bankier, R. Bechhofer, C. A. Bennett, J. Berkson, C. A. Bicking, C. I. Bliss, I. Blumen, R. C. Bose,

A. H. Bowker, R. A. Bradley, A. E. Brandt, R. Brickley, S. H. Brooks, R. W. Burgess, P. J. Burke, I. W. Burr, L. D. Calvin, A. G. Carlton, D. G. Chapman, A. Charnes, H. Chernoff, C. W. Churchman, A. G. Clark, W. G. Cochran, A. C. Cohen, Jr., E. L. Cox, Gertrude M. Cox, J. H. Curtiss, C. Daniel, G. B. Dantzig, R. B. Dawson, Jr., B. B. Day, A. De la Garza, F. Del Priore, D. B. DeLury, W. E. Deming, L. Derrick, J. L. Dobby, T. Donnelly, J. E. Dowd, J. R. Duffett, A. J. Duncan, D. B. Duncan, C. W. Dunnett, G. L. Edgett, C. Eisenhart, B. Epstein, J. W. Fertig, D. Fraser, J. E. Freund, Kathryn Froelich, L. Gerende, R. Goodman, C. H. Goulden, M. H. Gourary, F. Graybill, B. G. Greenberg, S. W. Greenhouse, L. Gunlogson, P. Gunther, M. Gurney, I. Guttman, R. K. Haddad, R. J. Hader, K. W. Halbert, M. Halperin, J. F. Hannan, M. H. Hansen, G. W. Brier, G. M. Harrington, B. Harris, B. Harshbarger, H. L. Harter, L. H. Herbach, G. R. Herd, W. Hoeffding, R. G. Hoffman, P. G. Homeyer, W. C. Hood, L. H. Hook, W. H. Horton, H. Hotelling, E. E. Houseman, C. J. Hoyt, J. F. Hudson, J. S. Hunter, P. Irick, J. E. Jackson, Carol M. Jaeger, H. L. Jones, M. Kac, L. Katz, Harriet J. Kelly, N. Keyfitz, A. W. Kimball, B. F. Kimball, E. P. King, C. J. Kirchen, L. Kish, K. H. Kramer, W. Kruskal, R. B. Ladd, R. O. Laine, T. A. Lamke, F. C. Leone, R. Lessard, G. J. Lieberman, R. Likert, B. Lipstein, S. B. Littauer, G. F. Lunger, C. J. Maloney, J. Mandel, E. S. Marks, R. H. Matthias, J. W. Mauchly, P. Meier, Margaret Merrell, W. J. Merrill, H. A. Meyer, R. Mirsky, S. Monro, C. B. Moore, R. Moore, M. Morrison, N. Morse, J. Moshman, G. E. McCreary, Horace W. Norton, James A. Norton, Jr., G. E. Noether, F. G. Cornell, G. E. Nicholson, Jr., Ingram Olkin, E. G. Olds, Robert E. Odeh, W. R. Pabst, Jr., Boyd Z. Palmer, W. E. Patte, A. E. Paull, E. W. Pike, Howard Raiffa, Mrs. L. K. Randolph, J. S. Rhodes, Robert Roeloffs, John H. Roseboom, H. M. Rosenblatt, Joan R. Rosenblatt, Murray Rosenblatt, Irving Roshwalb, S. N. Roy, Jacques St. Pierre, Daniel E. Sands, A. E. Sarhan, F. E. Satterthwaite, Henry Scheffé, Marvin Schneiderman, H. L. Seal, Daniel Seigel, Richard H. Shaw, Loren R. Sheldahl, David Sheppard, Walt R. Simmons, H. F. Smith, Walter Smuk, Herbert Solomon, Paul N. Somerville, F. F. Stephan, David S. Stoller, Hale C. Sweeney, Nancy Symons, Donovan J. Thompson, W. A. Thompson, Jr., William R. Thompson, Leo J. Tick, John W. Tukey, Charles R. M. Tuttle, G. W. Tyler, W. Vander Byl, D. F. Votaw, Jr., A. J. Wadman, John R. Walter, Lionel Weiss, Samuel Weiss, John S. White, Alfred G. Whitney, William Wolman, Max A. Woodbury, R. Wormleighton, A. W. Wortham, W. J. Youden, Samuel Zahl, Marvin Zelen, George Zyskind.

The Program was as follows:

FRIDAY, SEPTEMBER 10, 1954

10:00 a.m. Theory of Sampling Fish Populations (Cosponsored by A.S.A. and Biometric Society)

Chairman: D. B. DeLury, Ontario Research Foundation
Papers: 1. *Combined Estimation Methods in Sampling Fish Populations*, D. G. Chapman, University of Washington
2. *Some Admissible Tag-Recapture Procedures*, D. S. Robson, Cornell University
Discussion: E. L. Cox, Case Institute of Technology

3:00 p.m. The Joint Effects of Reading Errors and Grouping on Standard Methods of Statistical Inference. (Invited address). (Cosponsored by A.S.A. and Biometric Society)

Chairman: R. L. Anderson, North Carolina State College
Speaker: Churchill Eisenhart, National Bureau of Standards

4:00 p.m. Multiple Comparison and Multiple Decision Procedures. (Co-sponsored by A.S.A. and Biometric Society)

Chairman: John W. Tukey, Princeton University

Papers:
1. *A Survey of Multiple Comparison Procedures*, Jerome Cornfield, National Institute of Health
2. *A Survey of Multiple Decision Procedures*, Robert Bechhofer, Cornell University

Discussion: R. C. Bose, North Carolina State College and W. G. Cochran, Johns Hopkins University

7:15 p.m. Council Meeting

SATURDAY, SEPTEMBER 11, 1954

10:30 a.m. Invited Addresses (Cosponsored by Econometric Society)

Chairman: Howard Raiffa, Columbia University

Papers:
1. *Estimation of the Components of Stochastic Structures*, J. Wolfowitz, Cornell University
2. *Nonparametric Large Sample Theory*, Wassily Hoeffding, University of North Carolina. Special invited address.

2:00 p.m. Contributed Papers I

Chairman: G. Lieberman, Stanford University

Papers:
1. *On Quadratic Estimates of Variance Components in Balanced Models*, A. W. Wortham, Chance Vought Aircraft and Oklahoma A and M College
2. *The coefficients in the Best Linear Estimate of the Mean in Symmetric Populations*, A. E. Sarhan, University of North Carolina
3. *Distribution of Linear Contrasts of Order Statistics*, Jacques St. Pierre, University of North Carolina
4. *Note on Fourier Periodogram Analyses of Time Series*, B. F. Kimball, New York State Public Service Commission
5. *Univariate Two-Population Distribution-Free Discrimination*, David S. Stoller, Rand Corporation
6. *New Types of Easily Constructed Partially Balanced Incomplete Block Designs*, John Mandel and Marvin Zelen, National Bureau of Standards
7. *The Stochastic Convergence of a Function of Sample Successive Differences*, Lionel Weiss, University of Virginia
8. *On a Modified T^2 Problem*, Ingram Olkin, Michigan State College and S. S. Shrikhande, College of Science, Nagpur
9. *The Validity of Sheppard's Corrections for Grouping*, F. J. Anscombe, University of Cambridge and Princeton University
10. *Unbiased Tests Based on Unbiased Estimators*, Reed B. Dawson, Department of Defense
11. *The Mean Square Error of the Sample Median*, Harold Hotelling, University of North Carolina
12. *The Moments of the Sample Median*, J. T. Chu and Harold Hotelling, University of North Carolina
13. *Distribution of the Largest Vote in Unstructured Random Balloting*, Leo Katz, Michigan State College

14. *Statistical Programming*, D. F. Votaw, Jr., Yale University
15. *Exact Tests of Significance for Combining Inter- and Intra-Block Information in Incomplete Block Designs (Preliminary Report)* (By title), Marvin Zelen, National Bureau of Standards
16. *Moments and Related Quantities of the Null Distribution of Linear Contrasts of Order Statistics in the Case of Three Populations* (By title), Jacques St. Pierre, University of North Carolina.
17. *Application of Faà di Bruno's Formula in Mathematical Statistics* (By title), Eugene Lukacs, Office of Naval Research.
18. *On Simultaneous Minimax Point Estimation* (By title), Waldo A. Vezeau and Koichi Ito, St. Louis University
19. *Estimation of Structural Parameters when the Number of Incidental Parameters is Unbounded* (By title), J. Wolfowitz, Cornell University
20. *On Power Properties of Certain Simultaneous Tests* (By title), K. V. Ramachandran, University of North Carolina.
21. *On Tests of Normality and Other Tests of Goodness of Fit Based on Distance Methods* (By title), M. Kac, J. Kiefer, and J. Wolfowitz, Cornell University
22. *Tolerance Regions* (By title), D. A. S. Fraser and I. Guttman, University of Toronto.

SUNDAY, SEPTEMBER 12, 1954

12:00 noon. Council Meeting

2:00 p.m. Contributed Papers II

Chairman: Roger Lessard, Ecole Polytechnique de Montreal

Papers:
1. *Comparison of the Power of Nonparametric Two Sample Tests against Normal Alternatives*, Benjamin Epstein, Wayne University
2. *On the Distribution of Radial Errors Having Normally Distributed Components*, A. C. Cohen, Jr., University of Georgia
3. *Confidence Bounds on Departures from a Particular Kind of Multi-Collinearity of Means*, S. N. Roy, University of North Carolina
4. *The Efficiency of Tests*, Wassily Hoeffding and Joan R. Rosenblatt, University of North Carolina
5. *On a Decision Procedure to Select the Population with the Largest Mean (Preliminary Report)*, R. C. Bose and Jacques St. Pierre, University of North Carolina
6. *Most Economical Multiple-Decision Rules* (By title), William Jackson Hall, University of North Carolina

4:00 p.m. Business Meeting

MONDAY, SEPTEMBER 13, 1954

8:30 a.m. Computing Techniques (Cosponsored by Econometric Society)

Chairman: George B. Dantzig, The Rand Coporation

Papers:
1. *Linear Programming on the I.B.M. 701*, Kurt Eisemann, International Business Machines Corp.
2. *Recent Extensions and Revisions of the Simplex Method*, William Orchard-Hays, The Rand Corporation
3. *Separable Convex Functionals and Generalized Simplex Methods*, C. E. Lemke and A. Charnes, Carnegie Institute of Technology

Discussion: Alan J. Hoffman, U.S. Department of Commerce

10:30 a.m. Biological Cycles (Cosponsored by A.S.A., Biometric Society, and Econometric Society)

Chairman: J. W. Hopkins, National Research Council (Canada)
Papers: 1. *Population Dynamics*, D. A. MacLulich, Royal Canadain Air Force
2. *Statistical Problems and Techniques in Population Cycle Analysis*, Mark Kac, Cornell University
Discussion: D. B. DeLury, Ontario Research Foundation

LIONEL WEISS
Associate Secretary

MINUTES OF MEMBERSHIP MEETING, SEPTEMBER 12, 1954

A business meeting was called to order at 4:10 P.M., September 12, 1954 in the Sheraton Mount Royal Hotel, Montreal, P.Q. by President E. G. Olds. About 55 members were present.

The Secretary read the minutes of the Annual Meeting of December 29, 1953.[1] The minutes were approved as read.

The Secretary moved, on behalf of 43 members of the Institute, the adoption of the following amendment to the Bylaws:

"All meetings of the Institute shall be held on the basis of no racial segregation. In particular, prior to determining the place of a forthcoming meeting the Secretary of the Institute shall ascertain that meeting halls, eating facilities, and housing accommodations adequate for the expected attendance shall be available on a nonsegregated basis, and that all social events connected with the meeting shall be nonsegregated."

The Secretary then moved, on behalf of 31 members of the Institute, the substitution of the following amendment for the one given immediately above:

"It is the policy of the Institute of Mathematical Statistics that all its meetings shall be held on a completely nonsegregated basis. In particular, prior to determining the place of a forthcoming meeting, the Secretary of the Institute of Mathematical Statistics shall ascertain that meeting halls and eating facilities adequate for the expected attendance will be available on a nonsegregated basis and that all social events connected with the meetings shall be nonsegregated. Every effort shall be made to provide nonsegregated housing accommodations consistent with the laws of the locality of a forthcoming meeting."

The President announced that members present would be asked to vote on the same ballot forms as used by those voting by mail and called for discussion of both questions. There was no discussion. Ballots were collected and the tellers retired to count the ballots. After adjournment of the meeting the results of the balloting were posted as follows: 575 ballots were cast. On the question of substituting the second amendment above for the first, 126 voted "yes," 437 voted "no," 12 ballots contained no vote on the question. The next vote was therefore on the adoption of the first amendment above. On this question 311 voted "yes,"

[1] Published in the *Annals*, Vol. 25 (1954) pp. 191–193.

236 voted "no," 28 ballots contained no vote on the question. Lacking a two-thirds majority in favor, the motion lost.

The Secretary-Treasurer gave an informal report.

The Program Coordinator reported on the progress of plans for future meetings.

The President reported discussion in the Council of several progress reports of committees of the Institute. Members were told that the Committee to Explore the Desirability of Changing Time of Winter Meetings would welcome comments on whether or not they favored Christmas meetings and on other related questions.

The appointment of S. S. Wilks as representative of the Institute in the Division of Mathematics of the National Research Council for the term July 1954 to June 1957 was announced.

The election by the Council, on the nomination by a Committee to Nominate an Editor, of Erich L. Lehmann for the term 1956–1958 was announced.

The election by the Council, on nomination by the Editor, of H. E. Daniels as an Associate Editor for a term beginning immediately and continuing until December 1955 was announced.

The addition of T. W. Anderson, Jr., Z. W. Birnbaum, D. H. Blackwell, J. L. Doob, W. Feller, John Gurland, C. M. Stein, G. E. Nicholson, Milton Sobel, and Murray Rosenblatt to the Committee on a Summer Statistical Institute and the resignations from this committee of Gertrude M. Cox and Herbert Solomon were announced.

The addition of Arnold Court to the Committee to Reexamine the Constitution and Bylaws was announced.

The appointment of W. Feller as Rietz Lecturer for 1955 was announced.

A. H. Bowker was called on to report informally on the progress of discussions within the Committee on Activities and Development.

Boyd Harshbarger presented the following resolutions which were passed by the meeting.

"Whereas Professor Roger Lessard of the Local Arrangements Committee has provided excellent arrangements for this meeting be it resolved that we the members of the Institute of Mathematical Statistics express our appreciation and thanks to him. The Secretary will record this resolution as a permanent part of the minutes of the meeting and a copy will be sent to Professor Lessard.

"Whereas Dr. Edwin G. Olds, President, Dr. K. J. Arnold, Secretary, and Dr. Lionel Weiss, Associate Secretary, have shown outstanding leadership, be it resolved that we the members of the Institute of Mathematical Statistics express appreciation and thanks to them. The Secretary will record the resolution as a permanent part of the minutes of the meeting and copies will be sent to Dr. Olds, Dr. Arnold and Dr. Weiss.

"Whereas Dr. D. B. DeLury and his associates have arranged a stimulating and creative program be it resolved that we the members of the Institute of Mathematical Statistics express appreciation and thanks to them. The Secretary will record this resolution as a permanent part of the minutes of the meeting and a copy will be sent to Dr. DeLury.

"Whereas the City of Montreal and its officials have extended a cordial welcome and have given a gracious and generous reception for the members and guests, be it resolved that we

the members of the Institute of Mathematical Statistics express appreciation and thanks. The Secretary will record the resolution as a permanent part of the minutes of this meeting."

The meeting adjourned at 5:15 P.M.

K. J. Arnold
Secretary

PUBLICATIONS RECEIVED

"A System of National Accidents and Supporting Tables," United Nations, 1953, $.50.

Anuario Estadistica de España, Instituto Nacional de Estadistica, 1953, 999 pp. 50 pesatas.

Beaumont, Ross A. and Richard W. Ball, *Introduction to Modern Algebra and Matrix Theory*, Rinehart & Company, Inc., New York, 1954, $6.00.

Chevalley, Claude C., *The Algebraic Theory of Spinors*, Columbia University Press, New York, 1954, $3.75.

"Concepts and Definitions of Capital Formation," United Nations, 1953, $.25.

Sanderson, Fred H., *Methods of Crop Forecasting*, Harvard University Press, Cambridge, 1954, $5.00.

"Statistics of National Income and Expenditure," United Nations, 1953, $.60.

Studi De Economia E Statistica, Ser. I, Vol. II, 1953, Università Di Catania Anno Accademico 1951-52, 389 pp.

INSTITUTIONAL MEMBERS

International Business Machines Corporation, New York
Iowa State College, Statistical Laboratory, Ames, Iowa
Michigan State College, Department of Mathematics, East Lansing, Michigan
Princeton University, Department of Mathematics, Section of Mathematical Statistics, Princeton, New Jersey
Purdue University Libraries, Lafayette, Indiana
Raytheon Manufacturing Company, Newton, Massachusetts
State University of Iowa, Iowa City, Iowa
University of California, Statistical Laboratory, Berkeley, California
University of Illinois, Urbana, Illinois
University of North Carolina, Institute of Statistics, Chapel Hill, North Carolina
University of Washington, Laboratory of Statistical Research, Seattle, Washington